DOCUMENT RESUME

ED 045 152 52 LI 002 369

TITLE	Seminar on Foci for Progress in Scientific Publications: A Summary and Final Report.
INSTITUTION	Battelle Memorial Inst., Columbus, Ohio.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-0-0074
PUB DATE	[70]
CONTRACT	OEC-0-70-1493(506)
NOTE	116p.: Seminar held at National Science Foundation, Washington, D.C., Nov. 4, 1969
EDRS PRICE	EDRS Price MF-$0.50 HC-$5.90
DESCRIPTORS	*Information Dissemination, Information Needs, *Information Processing, *Information Sources, *Information Utilization, Scholarly Journals, Technological Advancement
IDENTIFIERS	*Scientific Publications

ABSTRACT

A one-day seminar was convened on November 4, 1969, to consider problems involved in communicating scientific information and the progress that has been made in improving the scientific publication mechanism. Thirty participants representing different types of information producers, users, and organizations contributed to the seminar discussion. Technologies that have facilitated machine-readable data bases were found to have materially facilitated the preparation of a variety of information products that are needed for both large collection and special interest group purposes. Additional research and lower costs are considered necessary to permit the extensive exploitation of time-shared, computer-based networks, but current hardware capabilities indicate that direct user involvement at individually accessible consoles is possible and is expected to materially aid the development of well-organized, user-oriented information bases. (Author)

BR-0-0074
PA 52

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

FINAL REPORT
Project No. 0-0074
Contract No. OEC-0-70-1493(506)

ED0 45152

SEMINAR ON FOCI FOR PROGRESS IN SCIENTIFIC PUBLICATIONS:

A Summary Report

Seminar held at:

National Science Foundation
Washington, D. C.

November 4, 1969

Hosted by:

Office of Science and Technology
Executive Office of the President

General Chairman:

Mrs. R. W. Swanson
U. S. Air Force Office of Scientific Research
Directorate of Information Sciences
Arlington, Virginia

Arrangements and Publications:

Battelle Memorial Institute
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

Although the Seminar summarized was sponsored by a contract with the Office of Education, U. S. Department of Health, Education, and Welfare, points of view or opinions stated do not necessarily represent official Office of Education position or policy.

LI 002 369

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Table of Contents

Page

Study Objective 1

Scope of the Seminar 2

Issues Regarding Scientific Publication 3

Summary 9

Appendixes

  I.--List of Participants 13

  II.--Contributors' Summaries 19

STUDY OBJECTIVE

Questions of how to improve the handling of the burgeoning literature of science and technology, and how to ensure the speedy communication of information about research and development (R & D) results to those that need them, are receiving increasing attention in many quarters. These questions have immediate importance to those who use, or should be taking advantage of, current information in their work. They are also important to R & D planners, managers, and budget officers whose projections are indirectly affected by the availability of information and the costs in time and manpower, as well as dollars, that must be incurred in its acquisition. Publishers and information service suppliers have been seeking and developing new approaches to packaging and disseminating information not only to gain new markets but also to alleviate mounting production costs.

Moreover, how information is organized, packaged, and used has implications both to education and to the exploitation of new media and technologies for educational purposes. Collecting and organizing information so that it can be conveniently and effectively used by people is a goal of educators as well as information processors. The proper use of information reflects education on the part of the user. Thus, information products should, optimally, be designed to educate as well as inform.

To consider ways of attaining these goals, a one-day seminar was convened that was hosted by the Office of Science and Technology, Executive Office of the President, and sponsored by the Library and Information Science Research Program, Bureau of Libraries and Educational Technology, U. S. Office of Education. Participants were invited who represented a cross-section of organizations concerned with matters of scientific publication, communication, and dissemination. These organizations included government sponsors of information handling activities, not-for-profit producers of information products, academicians concerned with uses of information in connection with economic and industrial development, information system designers, and scientist producers and users of information.

SCOPE OF THE SEMINAR

Thirty persons met on November 4, 1969, in the National Science Foundation Board Room to consider questions concerned with scientific publication. A list of the participants is given in Appendix I. Discussion proceeded according to the following agenda:

Session A. At this session, participants introduced the activities with which they are engaged. These activities are summarized in brief descriptions that are incorporated in Appendix II.

Session B. This session considered the relevance of current work to conjectures of future needs at national and international levels. Particular emphasis was given to influences of current progress in science, technology, and education to likely future developments.

Session C. This session considered current activities critically in relation to perceived gaps in technologies, and possible effects of technological developments on future information activities.

Session D. This session elicited suggestions concerning the roles and missions that various organizations have and should have to foster the handling and communication of information. Public and private organizations, for-profit and not-for-profit, were involved in the participants' considerations.

ISSUES REGARDING SCIENTIFIC PUBLICATION

Lost--One-Third of the World's Scientific Literature is the title of a National Science Foundation brochure prepared in 1960.1/ This title dramatically encapsulates the dilemma resulting from the growing volume of scientific activity and technological development that have been fostered worldwide, particularly since the close of World War II. This booklet was primarily addressed to the language barrier that was estimated to cut about 95 percent of U. S. scientists off from nearly 30 percent of the world's literature. Less than 2 percent of scientists had reported to the National Science Foundation that they could read Russian, and reading capability by more than 1 percent of the scientific population was reported for only four other languages (German, French, Spanish, and Italian).

In addition to promoting the translation of foreign-language materials, this brochure recommended the voluntary contribution by scientific societies, scientific institutions, and Government agencies that conduct translation programs of translations made by them to national collection centers for centralized announcement and redistribution. The brochure also advocates the preparation of critical reviews and state-of-the-art papers for communicating information about foreign science.

The search for alternatives to the scientific periodical as a means of communicating information has been discussed in the literature for about forty years.2/ The practice of announcing the availability of papers through abstracts or notes in periodicals was introduced over a century ago in the Royal Society Proceedings. A useful analysis of the pros and cons of the scientific periodical was made in 1960 by Ralph Phelps and John Herling of the Engineering Societies Library. Among deficiencies, they include:

a. delays of a year or more between the time of acceptance of a paper and its appearance in print,

b. restrictions on the length of papers and other editorial demands that exclude historical and theoretical material and, sometimes, references,

c. a scattering of papers across a multiplicity of journals that hampers scientists from keeping informed and libraries from comprehensive coverage of a field,

---

1/ *Lost--One-Third of the World's Scientific Literature.* Washington, D. C., National Science Foundation, 1960. Report NSF 60-20.

2/ Ralph R. Phelps and John P. Herling, "Alternatives to the Scientific Periodical." *Unesco Bulletin for Libraries*, Vol. 14, No. 2, 1960, pp. 61-75.

d. poor matches between journal coverage and scientists' specializations that limit their interest to only a few papers in any periodical,

e. the high cost of journals to both individuals and libraries that limit acquisition, and, thereby, coverage, even by abstract services, and

f. wasteful incursions that the system imposes on the time of both the scientist-searcher for information and the scientist-reviewer and editor (the latter often on a volunteer basis that distracts from work in which he should be engaged).

Proposals for overcoming these deficiencies have included the following suggestions:

a. substitution of the individual paper for the periodical as the primary unit of distribution (the most often-made suggestion),

b. establishment of information centers on national and international levels for centralized publishing, bibliography, reference, copy distribution services, announcement and abstract periodicals, etc.,

c. selective dissemination of papers automatically to libraries and individuals in accordance with registration-of-interest lists (as well as on an as-requested basis),

d. establishment of an efficient method of weeding out poorly written and unoriginal papers before they are published, possibly by a central editorial bureau of experts,

e. coupon systems for ordering papers from abstract journals,

f. production of abstract journals that describe papers in all languages in a common language,

g. issuance of tables of contents of periodicals from which desired papers could be ordered (proposed by J. D. Bernal in the late 1940's and first attempted by E. Garfield in 1953), and

h. simultaneous preparation by authors of a full report, a two-page abridgement, and an abstract, selected papers and abstracts being published in journals with all abstracts sent to journal subscribers (possibly on fileable cards) who could then order unpublished papers of their choice.

Phelps and Herling contend that arguments against the periodical do not overcome the following criticisms of alternative proposals:

a. the tendency of more limited distribution to impose restrictions on scientists' reading, thereby reducing the value browsing and casual reading are believed to have on creative thinking and education,

b. the threat of censorship inherent in centralized regulation of publication and dissemination,

c. the administrative burden on distributing agencies and on documentation centers and libraries attendant on "separates" schemes,

d. higher processing costs that could no longer be at least partially defrayed by the financial support of advertisers,

e. loss of the skills of scientific editors who would not be expected to serve centralized agencies,

f. loss of intrinsic merits of journal publication and the associated confidence readerships have in those that have gained reputations for merit, particular scientific interests, etc.,

g. damage to the coherence and vitality of many professional societies whose raison d'être, at least in part, is their periodicals, and

h. the minor role that publication is presumed to have in scientific communication, much of this having been shown to be effected through interpersonal exchange, informal correspondence, attendance at meetings, etc.

Although problems of scientific publication and communication of information are often considered from either the publisher's or the user's point of view, too often the role of the originator or producer of information is overlooked. Another NSF brochure contains the following admonition:

> "The scientific information problem begins in the research and development establishments where project results are converted into documentary form. These information sources include Government agencies and their contractors, private industry, colleges and universities, and independent research institutions. Not all users are sources, but every source is, of necessity, a user of scientific information."3/

3/ *Some Problems in the Reporting of Research Results*. Washington, D. C., National Science Foundation, circa 1960.

Some of the activities suggested to these source-users to promote timely and meaningful communication include:

a. prompt dissemination of information on current research in progress,

b. better program planning and selection of papers presented at symposia and conferences,

c. exploitation of new media and formats to speed the information flow and simplify bibliographic, storage, and retrieval phases,

d. the development of new techniques for separating and consolidating information to facilitate retrieval and use of information,

e. reporting of negative results,

f. use of and improvement in the quality of abstracts accompanying papers,

g. source indexing to guide and assist indexers of periodical and documentation services,

h. the development of uniform terminology and coding systems to facilitate bibliographic processing, storage, and retrieval,

i. use of conventional publications that provide the best means of making information readily available,

j. prompt announcement of industrial reports and university theses with provisions for making them available, and

k. reassessment of policies and criteria governing the application and removal of security and other limitations on access to the literature to facilitate information flow.

On the international scene, several organizations have focused on problems of scientific publication and information flow. These include the International Federation for Documentation (FID), the International Federation of Library Associations (IFLA), the International Council of Scientific Unions (ICSU) Abstracting Board (ICSU/A[illegible]) and Committee on Data for Science and Technology (CODATA), the UNESCO International Advisory Committee on Documentation, Libraries, and Archives, the Technical Information Panel of AGARD (NATO's Advisory Group for Aerospace Research and Development), and UNISIST (a body created by ICSU and UNESCO to examine the feasibility of a world science information system).

UNESCO Working Parties have, for example, been considering such topics as:

a. relationships between abstracting, indexing, and primary publications, including the relative value of indexes and abstracts for users,

b. the role of scientific periodicals and alternatives as ways of publishing scientific papers,

c. the organization and functioning of abstracting services in various branches of science and technology,

d. a code of good practice for scientific publications,

e. standardization for primary and secondary publications of such elements as abbreviations of titles of periodicals, transcription of cyrillic characters, and bibliographic citations,

f. the creation of regional, subject, and national associations of editors of scientific and technical periodicals, and

g. improvements in the preparation and availability of publications resulting from scientific meetings.

The contents of Appendix II reflect various approaches that have been and are being attempted that emanate, at least in part, from the deliberations reviewed above. Participants at the Seminar are finding that machine-readable data bases are facilitating the production of a variety of information products for both archival and large central-store purposes and for the needs of special interest groups. Analytic products in the nature of evaluative reports and states-of-the-art that are being developed by commercial organizations are proving salable with the proviso that a good marketing mechanism is a necessity for informing potential users of the products. Emphasis was suggested on user involvement to ensure the intelligent use of information products and information. Technologies are making it possible to give users direct access to large central files and to enable them to restructure files in accordance with their particular needs from consoles that can become part of their personal office equipment. Machine-based networks are expected to have the effect of bringing members of common-interest communities who are geographically dispersed in closer contact through the medium of their consoles. File content improvement and enhancement are anticipated as users gain the ability to add comments and corrections to stored material. Mechanization is also expected to enable quantitative assessment of use that is made of information and up-grading of information stores through monitoring programs that need not interfere with monitored users. To date, most of the information products that have been developed have been innovations of producers. The greater user involvement that is anticipated is expected to stimulate innovations from users. Increased investigation of television, audio, and audio-visual technologies was suggested for the kinds of information they are well suited to transmit. Different types of information packages with the same information content were not viewed as detrimentally duplicative when each served a useful and needed purpose. Consideration of possibilities of cost sharing by various organizations was recommended to spread the burden of admittedly high costs of establishing and maintaining needed information bases.

SUMMARY

A one-day seminar was convened on November 4, 1969, to consider problems involved in communicating scientific information and the progress that has been made in improving the scientific publication mechanism. Thirty participants representing different types of information producers, users, and organizations contributed to the seminar discussion. Technologies that have facilitated machine-readable data bases were found to have materially facilitated the preparation of a variety of information products that are needed for both large-collection and special-interest-group purposes. Additional research and lower costs are considered necessary to permit the extensive exploitation of time-shared, computer-based networks, but current hardware capabilities indicate that direct user involvement at individually accessible consoles is possible and is expected to materially aid the development of well-organized, user-oriented information bases.

appendixes

APPENDIX I.--LIST OF PARTICIPANTS

Mr. Scott Adams
National Academy of Sciences
Office of the Foreign Secretary
C/O Dr. Harrison Brown
2101 Constitution Avenue, N. W.
Washington, D. C. 20418
(202) 393-8100

Dr. Burton W. Adkinson
National Science Foundation
Office of Science Information
Service
Washington, D. C. 20550
(202) 632-5824

Colonel Andrew A. Aines
Executive Office of the President
Office of Science and Technology
Washington, D. C. 20506
(202) 395-3547

Dr. Jordan J. Baruch
EDUCOM
100 Charles River Plaza
Boston, Massachusetts 02114
(617) 227-1805

Mr. Lawrence H. Berul
Auerbach Corporation
121 North Broad Street
Philadelphia, Pennsylvania 19107

Mr. Ralph L. Bisco
Council of Social Science
Data Archives
University of Pittsburgh
Social Science Building, Room 213
Pittsburgh, Pennsylvania 15213
(412) 621-3500, Ext. 352

Mr. Lawrence F. Buckland
Inforonics, Inc.
146 Main Street
Maynard, Massachusetts 01754
(617) 897-8815

Dr. Lee G. Burchinal
U. S. Office of Education
Bureau of Research
Division of Information Technology
and Dissemination
Washington, D. C. 20202
(202) 962-6346

Mr. David W. Cardwell
Oak Ridge National Laboratory
Reactor Division
P. O. Box Y
Oak Ridge, Tennessee 37830
(615) 483-8611, Ext. 35647

Mr. Ralph L. Darby
Battelle Memorial Institute
Information Operations Division
505 King Avenue
Columbus, Ohio 43201
(614) 299-3151, Ext. 2071

Prof. Martin Davidson
North Texas State University
Department of Economics
Denton, Texas 76203
(817) 387-4511

Mr. Bernard K. Dennis
Battelle Memorial Institute
Information Research Section
1755 Massachusetts Avenue, N. W.
Washington, D. C. 20036

Dr. Richard A. Dershimer
American Education Research
Association
1201 Sixteenth Street, N. W.
Washington, D. C. 20036
(202) 223-9485

Dr. James L. Dolby
R&D Consultants Company
885 No. San Antonio Road
Los Altos, California 94022
(415) 948-0370

Prof. John S. Gilmore
Denver Research Institute
Industrial Economics Division
University of Denver
Denver, Colorado 80210
(303) 753-3207

Mr. John R. Hart
North Carolina State University
Industrial Extension Service
P. O. Box 5506
Raleigh, North Carolina 27607
(919) 755-2357

Prof. Jack Heller
State University of N. Y. at
Stony Brook
Dept. of Applied Analysis
Stony Brook, Long Island, N. Y.
(516) 246-6773

Mr. Conyers Herring
Bell Telephone Labs., Inc.
Murray Hill, New Jersey 07911
(201) 582-3000, Ext. 2308

Dr. Richard L. Kenyon
American Chemical Society
1155 Sixteenth Street, N. W.
Room 604A
Washington, D. C. 20036
(202) 737-3337, Ext. 315

Dr. H. William Koch
American Institute of Physics, Inc.
335 East 45th Street
New York, New York 10017
(212) 685-1940

Prof. Melvin Kranzberg
Case Western Reserve University
Crawford Building
Cleveland, Ohio 44106
(216) 368-4114

Mr. Robert Landau
C/O Executive Office of the President
Office of Science and Technology
Washington, D. C. 20506
(202) 395-3547

Mr. Jeffrey Norton
CCM Information Sciences, Inc.
909 Third Avenue
New York, New York 10022
(212) 935-3080

Dr. Sidney Passman
U. S. Arms Control & Disarmament
Agency
U. S. Department of State
2201 C Street, N. W.
Washington, D. C. 20451
(202) 632-8079

Mr. S. A. Rossmassler
Executive Office of the President
Office of Science and Technology
Washington, D. C. 20506
(202) 395-3547

Prof. Otto H. Schmitt
University of Minnesota
Department of Biophysics
Minneapolis, Minnesota 55455
(612) 373-3345

Miss Louise Schultz
BioSciences Information Service
2100 Arch Street
Philadelphia, Pennsylvania 19103
(215) 569-1100

Mr. Jack Sieburg
Office of the Judge Advocate
General USAF (AFJACIL)
LITE Activity
4710 Holly Street
Denver, Colorado 80210
(303) 825-1161, Ext. 344

Dr. Philip Siegmann
American Psychological Association
1200 Seventeenth Street, N. W.
Washington, D. C. 20036
(202) 296-7310, Ext. 219

Mr. Charles H. Stevens
Massachusetts Institute of Technology
Project Intrex
Cambridge, Massachusetts 02139
(617) 864-6900, Ext. 6051

Mrs. Rowena W. Swanson
Air Force Office of Scientific Research
Directorate of Information Sciences
1400 Wilson Boulevard
Arlington, Virginia 22209
(202) 694-5407

Dr. Charles Thompson
Northwestern University
Department of Industrial Engineering
and Management Sciences
Evanston, Illinois 60201
(312) 492-3680

Dr. David Weeks
Biological Sciences Communication Project
George Washington University
2000 P Street, N. W.
Washington, D. C. 20006
(202) 462-5828

APPENDIX II.--CONTRIBUTORS' SUMMARIES

Contents Page

ADAMS, SCOTT
Repackaging: Retrospect and Prospect 21

ADKINSON, BURTON W.
OSIS Contributions to Progress in Scientific Publication 27

BERUL, LAWRENCE H.
Activities in Communicating Technical Information 33

BISCO, RALPH L.
Role of Computerized Archives for the Social Sciences and Humanities 37

CARDWELL, DAVID W.
New Approaches to Handling Technical Information 41

DARBY, RALPH L.
Knowledge-Transfer Tools of Scientists and Technologists 45

DAVIDSON, MARTIN J.
Information's Fate and Future in the Marketplace 51

DERSHIMER, RICHARD A.
Activities to Enhance Communication in the AERA 55

DOLBY, JAMES L.
Developing Software Interfaces for Information Processing ... 59

GILMORE, JOHN S.
The User of Technological Information, Target or Participant? 63

HART, JOHN R.
Information Services for Industry 69

HELLER, JACK
Museum Data as an Information Resource 73

HERRING, CONYERS
Current Economics of Primary Journals in Basic Science and Technology 77

KENYON, RICHARD L.
Communicating Chemical Information 81

Appendix II Contents Continued Page

KOCH, H. WILLIAM
User Orientation of Primary Journals in Physics 87

KRANZBERG, MELVIN
The Technology Transfer Analog 93

NORTON, JEFFREY
Commercial Information Products and Services 99

PASSMAN, SIDNEY
Some Considerations Concerning Impedances in the Scientific Communication Network 105

SIEBURG, JACK
Information Potentials of Full-Text Computerized Records ... 109

SIEGMANN, PHILIP J.
Exploitation of Files for Information Products 113

STEVENS, CHARLES H.
Improving User Access to Information Through Applications of Technology 117

SWANSON, ROWENA W.
Proposals for Enhancing the Handling of Information 121

THOMPSON, CHARLES W. N.
Experiments with Responses of Users: Preliminary Findings 125

WEEKS, DAVID
The Role of Informal Communication 129

SCOTT ADAMS

National Academy of Sciences
Office of the Foreign Secretary
Washington, D. C.

REPACKAGING: RETROSPECT AND PROSPECT

A memorandum is reproduced below that has had an interesting history. Originally prepared by the author for Dr. F. B. Rogers to discuss at the Gordon Conference in the summer of 1961, it was shared with Dr. Alvin M. Weinberg on the occasion of the NFSAIS meeting in 1962, and subsequently with the Heller Associates as they were preparing the "Organization X" report. It has not been revised or updated in the eight intervening years. It was the philosophic base of the exploitation of the MEDLARS data base in cooperation with *need groups*. MEDLARS now produces for distribution by others 15 specialized packages from a common data base.

Viewed in retrospect, sponsors of mission-oriented abstracting and indexing publications have not so far broken down the doors of the custodians of the discipline-oriented data bases. The services themselves have experimented with repackaging their own products. The work on compatibility and convertibility of outputs which would enable others to do this did not get done.

Viewed in prospect, it is the group of secondary processors of tapes for SDI purposes who are accomplishing the function of melding the products of the large discipline-based systems. This group has made rapid progress in recent years. The group is international, and its work will have considerable impact on the future provision of information services. Whether it will concern itself with the generation of published products remains to be seen. It appears more likely that the information industry will fill this need.

DISCIPLINE AND MISSION ORIENTED SCIENCE IN RELATION TO THE ABSTRACTING PROBLEM

"Indexing and abstracting services traditionally reflect the organization of the sciences which have created, supported, and used them. Thus *Chemical Abstracts* has reflected scientific assumptions 1) that there is a disciplinary area of science and technology with a common concern for the constitution and reconstitution of matter and 2) that the practicioners, teachers, and researchers in this general *discipline* divide the field by agreed principles and organize laboratories, academic courses, and professional associations accordingly. History records that these considerations have conditioned the traditional organization of abstracting services--by scientific discipline.

Science is in the middle of a revolution today. While organization by discipline has survival value for education, the frontiers of research are characterized by categorical or mission orientation, where the knowledge and skills of multiple disciplines are brought to a common focus on a problem area. Conflicts between discipline- and mission-oriented research constitute a major issue in the planning and administration of research. Similarly, this conflict has a profound impact on the organization of information systems, the implications of which, possibly more far reaching than the sheer volume of scientific publication, or increased specialization, we are just beginning to understand.

Multidisciplinary research obtains in both basic and applied fields. It is characterized by *ad hoc* cooperation among the traditional disciplines toward solutions of problems considered socially important. This cooperation may or may not be enduring. Philosophically considered, multidisciplinary research may demonstrate the *seamless fabric of sciences*; in practice, it is concerned with borrowing threads from many areas of the fabric to reweave them into a new design.

The information requirements of multidisciplinary research are critical. Once a categorical problem has been identified and researchers from multiple disciplines assembled, two basic needs are expressed:

1. To pool the knowledge of the contributing disciplines which relates to the problem, as a fund of common knowledge is essential for a common understanding.

2. To provide a *current awareness* service--abstracts, indexes, newsbulletins, etc.--reflecting current contributions from all the scientific disciplines which bear on the problem.

The effect of this trend is marked. First, the requisite retrospective knowledge derives from tearing down disciplinary organized information to reconstitute it with the new mission orientation. Second, the current services either require current dependence on existing discipline-oriented services, or must duplicate their activities.

The problem becomes acute when the economic dimension is added. Mission-oriented research is by definition of social importance, and on the frontiers of sciences. It is relatively easy to find generous financial support from governmental or other sources. Such support, however, tends to decrease support available for the traditional discipline-oriented services, which have acknowledged survival values.

If this trend continues, the following consequences are inevitable:

1. The discipline-oriented services will be placed in a critical position.

2. The problem-oriented services will proliferate, and uncoordinate duplication of effort is inevitable.

3. The scientific disciplines dependent on traditional abstracts and indexes for a permanent scientific record, or for educational purposes, will be the losers.

A radical rethinking of this problem is absolutely necessary. Supporting agencies have made tentative moves already in response to economic pressures, but long range planning is absent. This planning should be in the direction of coordination and economic inter-dependence of disciplinary and problem-oriented systems.

The fundamental requirements for a combination system which would 1) maintain discipline-oriented systems for education and record purposes, and 2) permit the economic establishment of multiple mission-oriented systems as required, are:

1. Coverage of the scientific literature must be comprehensive.

2. This coverage must be both speedy and economical.

3. There must be maximum fluidity of the end product of such coverage, to enable it to be unitized and reassembled according to any mission-oriented need.

Computer technology, as it is applied to storage and retrieval of published information, provides a powerful instrument to bring this relationship about. The key requirement is that of unitizing information produced in disciplinary packages, manipulating it and producing it so that it may be applied for mission-oriented purposes--an operation computers are uniquely able to perform.

The following is a theoretical outline of a suggested solution:

1. Achieve an arbitrary division of the disciplines by supra-disciplinary groups (e.g. chemical sciences, physical sciences, earth sciences, biological sciences). Build up the existing services (*Chemical Abstracts*, *Biological Abstracts*, etc.), and create, under appropriate sponsorship, new ones to guarantee comprehensive coverage of world scientific publication.

2. Provide each such supra-disciplinary service with a computer facility. Work toward maximum compatibility of thesauri, and convertibility of output products.

3. Reverse the usual abstract/index relationship. Have these services prepare computer composed citation lists first, and superimpose abstracting operations as a secondary activity.

4. Encourage the mission-oriented services to rent screening services and listings from one or more of the supra-disciplinary services. These fee-for-service activities could, at the customer's option, also include:

a. full abstracts

b. photoduplication of all listed papers for later abstracting

c. contract preparation of complete mission-oriented abstract service.

Role of Government

While many exceptions may be found, the research programs of Federal agencies are predominantly mission-oriented. The Smithsonian Institution, which has survived as a home for much descriptive science, and the NSF are two exceptions to a list of agencies (DOD, NSF, HEW, NASA, AEC, etc.) which, generally speaking, all have categorical missions.

This is not to say that these agencies do not consciously support research by discipline, or that they are not involved in scientific and administrative problems incurred by trying to exist simultaneously in two worlds. It is to say that the resolution of their organizational and support problems are dictated by overriding forces representing subordination to goals which society considers important.

It follows that government sponsored adstracting services reflect the missions of the agencies sponsoring them, and it follows further that the sum-total of governmental scientific missions may always be less than the totality of science.

Therefore, if the proposed system is to be implemented through Federal planning and funds, it would appear necessary to seek authority at the level of the Office of Science and Technology.

Discussion

This proposal offers a possibility of integrating discipline- and mission-oriented information systems in such a manner as to:

1. Encourage the survival of the traditional services, by basing pay-as-you-go problem-oriented systems on their productivity.

2. Offer wide choice in the degree of dependence or integration, ranging from a citation screening service, to the supply of completed abstracts.

3. Effect marked economies in those operations which can best be done centrally--the collection and primary screening of the scientific literature. Present duplication of these necessary preliminary steps now constitutes a heavy expense to all concerned.

4. Satisfy both groups with promptly produced indexes or announcement services.

5. Satisfy both groups with comprehensive abstracting coverage.

6. Provide a basis for a wide variety of other forms of secondary publication: review papers, compendia, handbooks, bibliographies, etc.

This is a theoretical discussion of a complex practical problem; there are many difficulties to be overcome in further planning and in implementation. It represents, however, the philosophical underpinning of the MEDLARS project currently undertaken at the NLM. Granted success, it may be possible to test some of the theory in the not too distant future."

BURTON W. ADKINSON

National Science Foundation
Office of Science Information Service
Washington, D. C.

OSIS CONTRIBUTIONS TO PROGRESS IN SCIENTIFIC PUBLICATION

The objective of the Office of Science Information Service (OSIS) is to ensure that U. S. scientists and engineers have adequate information services to satisfy their needs. One approach in securing this objective has been the promotion of new and better techniques and systems for handling and disseminating science information; another is the improvement of existing systems and techniques to make them more effective. These approaches entail a wide variety of activities. For example, the OSIS encourages the major professional societies in the various scientific and engineering disciplines to integrate the information functions and services within their respective fields and to cooperate with their foreign counterparts and appropriate international organizations. At the same time, existing information services such as abstracting, indexing, and publishing are provided with partial support on a temporary basis. Universities and academic research organizations are encouraged to develop systems to meet the information requirements of scientists and engineers in their localities. Other projects we have sponsored include studies of communication patterns, the translation of current foreign-language journals, conferences, and travel to international meetings.

The OSIS does not itself operate any information service, reference service, or translation service. Its functions are limited to support of the activities of other organizations. In exercising this function, the OSIS is responsible for the prudent utilization of Federal (i.e., the taxpayers') funds. Through appropriate review mechanisms, the OSIS ensures that the activities supported by the funds appropriated for its programs contribute to the purposes and goals which describe the mission of the Office.

Problems with respect to scientific publication, or, more broadly, problems with respect to the communication of scientific and technical information, result to a considerable degree from a proliferation to the point that the scientific and technical community is now burdened with the wealth of published resources and seeks means to cope with the flood of new information embedded in the published record. The response to this set of problems by the scientific and technical community has resulted in innovative efforts and experiments of which a significant and seminal number have been supported by the OSIS.

Program Organization

At the present time, the organization of the OSIS reflects a program emphasis that ranges from research through development to operational support. Basic research and exploratory development are supported in our Research and Studies Program; the development of information systems is supported by the Information Systems Program; and operational support is provided by the Information Services Program. A separate unit was organized to administer support for the development of a national chemical information system, and a Special Projects Program has been given the responsibility of funding projects designed to improve university-centered information systems. This pattern of organization does not explicitly recognize the publication function as a discrete area for support. Other programs as well as those mentioned above find themselves concerned with publication support depending on whether the requirement is for deficit funding of an ongoing operation, for research in the ways publications are exploited or ways in which they might be improved, for developmental efforts when the requirements are known, and innumerable other conditions.

Patterns of Support for Innovative Scientific Publication

The patterns vary with the nature of the problem that is addressed. The problem set is conditioned by factors that are operative in the information-transfer environment. A simplistic enumeration might characterize these problems as follows:

1. Problems arising from the fact of the increasing amount of published materials competing for attention of the scientist;

2. Problems arising from economic relationships;

3. Problems arising from technological change;

4. Problems arising from changes in the total information transfer system.

This characterization tends to separate essentially interrelated problem situations, but appears to be useful for descriptive purposes. Thus, the problem of developing an innovative approach to the function of providing primary journal literature to scientists has been recognized in a number of differing contexts. One such approach assumes the availability of primary literature and explores the possibilities of improving the secondary information services designed to assist in informing about the availability, in organizing information about the availability, and in assisting the selection and retrieval process. This approach entails recognition that the secondary services are themselves issued in published form and that they require improvement in terms of comprehensiveness, speed, and reliability. Innovation in this area has involved exploitation of computer technology to expand the data base serving secondary services and to provide this data base in machine-readable form; it has involved the development of specialized and selectively oriented information services on the foundations of expanded and improved data bases; it has involved development of new products such as permuted title indexes, citation indexes, and selective dissemination services, which have become conventional in a surprisingly short time; and has led to developments which anticipate in some instances the displacement of traditional ink-print publication by on-line communication which nevertheless embodies basic publishing functions.

Examples of Innovative Efforts

The general problem emerging from the fact of increased publication involves the fact that the traditional journal seems to have become unwieldy as a mechanism for providing rapid and specific information needed by the individual scientist. A number of studies indicate problems with respect to the low percentage of readers of the individual articles, the increasing lag between completion of research and publication of findings, the economic burden on the individual subscriber in order to maintain an increasingly expensive dissemination mechanism, and other problems.

The response to these problems has emerged in a number of forms. The OSIS had received suggestions that journals might be packaged to the needs of the individual scientist. *Communications in Behavioral Biology* is an experimental project in which articles would be pre-indexed, reviewed by a staff of editors within two weeks of receipt, and processed for later computer printout or offset publication. Instead of appearing in a bound-journal format, the individual articles would be forwarded to subscribers on a flexible demand basis. The demand function is geared to the publication of abstracts of the indicated articles together with an index. The subscriber orders the desired articles, or indeed all articles, by virtue of the published abstracts and index. Since the initial formulation, the editors have arranged to secure abstracts of articles published in other journals, include these in the publication, and provide hard copies on demand. The experiment involves collaboration between the scientist-protagonist (a university professor), who carries editorial responsibility, and a commercial press. Both the university and the press contribute to the editorial expenses and publishing costs. The OSIS is a major contributor to the experiment, but foresees that successful experience will lead to a financially self-sufficient operation.

The American Mathematical Society has initiated a similar experiment with OSIS support. The *Mathematical Offprint Service* provides articles from a growing number of mathematical journals to individual scientists based on a combined notification and response function keyed to the interest profiles maintained and updated in the society's computer system. Interestingly enough, the index vocabulary which is basic to profile organization and management is presently being integrated with the indexing of the society's journals and is also a candidate thesaurus for wider adoption in mathematics publishing circles.

Economic problems beset the entire publishing industry. Scientific publication is particularly sensitive to economic pressures because of its essential nonprofit nature by virtue of the fact that scientific publishing is mainly carried on by professional societies and academic presses. The policy promulgated by the Federal Council for Science and Technology has guided the funding practices of the OSIS. Support has been provided to improve publishing practices, to pass through deficit periods, to improve and increase circulation, to exploit advanced technology, and improve overall management and editorial controls. Support for these purposes provided to ongoing operations is time limited. A condition of support is the availability of a plan that indicates how the protagonist, the editor or publisher, intends to manage without continuing subsidy and attain essential self-sufficiency within a three-year period.

The most significant technological advance affecting scientific publishing is the computer. Experimentation supported by the OSIS demonstrated the applicability of computers to journal production. The single most vexing problem in computer-controlled composition as applied to scientific publication relates to the capability of producing mathematical and scientific formulae via clerical keyboarding. Although this problem has not yet been solved in a manner as to produce the desired cost savings, developmental progress has been recorded in tying journal production via computerized composition to secondary bibliographic organization and dissemination. The American Chemical Society and the American Institute of Physics are implementing long range plans to harmonize the automated production of journals with the operation and dissemination of secondary abstracting and indexing services in both traditional printed and nontraditional tape formats.

The implications of the computer technology as applied to publication and dissemination of scientific information have not been and perhaps cannot be tested for some time to come. The scientific and technological community, especially that segment concerned with producing and publishing information as well as operating information services, including all sectors of our society, is faced with problems such as property rights in data, in copyright, and in software investments. These problems have barely surfaced and are not yet understood in their full complexity and dimension so as to permit an attack and effective solution efforts. The commission proposed by recently introduced legislation to study the effects of the computer technology on traditional copyright relationships; the efforts to study the effect of the computer on the rights of privacy of the individual; the economic consequences of developing communication networks which may make traditional publishing functions obsolete and dispose of traditional communication mechanisms such as the journal, the conference, and similar accepted communication media, are all signs of the current ferment to which the OSIS is sensitive. The funding of studies in these areas, and the cooperation with efforts to rationalize the steps required by these developments, is presently engaging the attention of the staff of the Office.

At the same time, conscious of its responsibilities to ensure the availability of scientific and technical information as assigned to it by the Congress, the OSIS continues to support the publications deemed necessary by the scientific and technical community consonant with its resources and program priorities.

LAWRENCE H. BERUL

Auerbach Corporation
Philadelphia, Pa.

ACTIVITIES IN COMMUNICATING TECHNICAL INFORMATION

I summarize here experiences we have accumulated as a producer and user of scientific publications.

Data Base Construction

We recently reported on an analytic study performed for the National Agricultural Library (NAL) that investigated the means by which NAL might interface its own automated programs with relevant bibliographic data extracted from other machine--readable data banks.(1) We surveyed nine data bases that represent almost seven million entries of potential value to NAL from 1959 to the present. We found that record formats vary significantly, variable-length fields being used most frequently to accommodate the variable formats of bibliographic data. Standardization of record formats is generally being regarded with a *wait-and-see* attitude; the Library of Congress' MARC II format has created the greatest amount of interest. With respect to data element representation, the CODEN for serial titles, used by four of the nine services, is the most widely accepted standard.

Generally, services in the profit-seeking sector appear to be evolving toward a service-center concept where the service will not only supply the data base but will also offer information services such as SDI and demand searching to the user. Those services in the nonprofit sector appear to be developing an opposite approach; they supply only a data base that the subscriber utilizes by whatever means are available to him. Packaged computer programs have not been made generally available, and users have had to develop their own search and processing software; three services do offer some programs to their users.

Though progress seems likely in the area of format conversion and restructuring, the problem of incompatibility among subject vocabularies still remains basically unsolved. One mode of data base interaction recommended to NAL was conversion of extracted portions of other data bases not only to NAL's own file structure, computer medium, and format, but also to its own indexing terminology, thereby permitting the merging of such data into NAL's own master bibliographic data base for searching and/or announcement. A study of the relative convertibility/compatibility of the vocabularies disclosed a total conversion rate for all vocabularies from 32.7 percent to 89 percent, based on a sample selection. The rates include an intellectual step of human conversion of terms that did not convert through a direct match or via associated terms. The vocabularies that most closely resembled the thesaurus format converted more successfully than uncontrolled vocabularies.

Auerbach Publications

Auerbach publications seek to provide information resources for the computer industry similar to those established by the West Publishing Co. for attorneys and by the Commerce Clearing House for accountants and attorneys. These publications include: Auerbach Standard EDP Reports, Auerbach Data Communication Reports, Auerbach Data Handling Reports.

These loose-leaf services are believed to go much further toward solving the information problem than abstracting and indexing services or most mechanized information retrieval systems. The key to their usefulness is the depth of information analysis and synthesis that is involved in their production. It should be noted that these services are offered in a field in which there is no paucity of printed matter. However, much of it is trivial, too poorly written, and lacking in a common organizational structure.

The Auerbach publications incorporate five basic features that contribute to their value to their user. These are the user's guide, the concept of performance measurement, the use of comparison charts, state-of-the-art reviews, and monthly or quarterly updating. The user's guide is a comprehensive outline that describes the format of the service. A faceted decimal numbering system is employed for rapid access and reference.

To facilitate user evaluation, a method of system performance measurement was devised that involves the use of standard benchmark problems. Each standard problem is fully detailed. For each computer configuration, for example, program timings are developed by an Auerbach analyst that are then incorporated in a chart for the service. Users need not have the same problem to exploit the benchmark evaluations.

Information Marketing

Perhaps the most important element for the survival of a high-cost information service is its approach to communicating information about itself that generates sales, i.e., marketing.

We have found that quality of a service does not attract customers of itself. It is necessary to actively promote the sale of information services. We have also found that services can be developed and maintained without government subsidies at prices that enable the financing of a reasonable in-house program of new product development.

We have, until now, relied primarily on the printed page for information transfer, and the loose-leaf method of updating. We are exploring products on microfilm and data bases in the form of magnetic tapes, disks, or responses via a time-sharing terminal. We are also considering other services, for example, a software evaluation reporting service, packaged training courses for EDP systems, a technology evaluation service, and new methods of packaging specialized libraries for the small and medium-size user.

Reference

1. H. B. Landau, *Research Study into the Effective Utilization of Machine-Readable Bibliographic Data Bases*. Final Report, No. 1582-100-TR-8, June 30, 1969. (Sponsored by the National Agricultural Library)

RALPH L. BISCO

University of Pittsburgh
Council of Social Science Data Archives
Pittsburgh, Pa.

ROLE OF COMPUTERIZED ARCHIVES FOR THE SOCIAL SCIENCES AND HUMANITIES

Social science data banks or data archives are a new facility for social science teaching and research. They developed out of the need by many social scientists to better exploit the masses of governmental, academic, and commercial data produced and accumulated over the past three decades. Their creation was made possible by the capabilities of modern computing technology. Although use of these new facilities has rarely equalled or exceeded the expectations of their creators, they have in their brief existence been found beneficial by thousands of graduate students and by hundreds of social scientists.

Basically, existing archives can be characterized according to the scope of their data collections and the geographic range of their services. There are four types of archives: 1) general-purpose, general-services archives such as the National Archives and Records Service or the Social Science Research Council Data Bank at the University of Essex, England; 2) special-purpose, general-service archives such as the International Data Service and Reference Library at the University of California, Berkeley, and the Roper Public Opinion Research Library at Williams College; 3) general-purpose, local-service archives such as the Data and Program Library Service at the University of Wisconsin and the Social Science Information Center at the University of Pittsburgh; and 4) special-purpose, local-service archives such as the Political Behavior Archives at the University of California, Los Angeles, and the Political Science Research Library at Yale University.(1)

Presently, most of the data stored in archives consists of copies of computer-usable collections prepared by various social research organizations worldwide. An increasingly significant proportion of holdings is beginning to consist of a variety of data--historical and administrative records, texts of publications, speeches, and news reports. Statistical data, heretofore available only in printed form, are being converted to computer-processable form by the archives themselves.

Development of archival facilities has necessarily been slow. First, support has had to be found to establish or continue them. Fees charged users for services cannot yet, and perhaps never will, make these organizations completely self-supporting. Data archives can be likened to libraries, cyclotrons, and medical research facilities whose benefit to society as a whole often requires more, or something other, than a strictly commercial justification for continuance.

Second, because these facilities are new and their requirements for staff are also new, they cannot usually recruit people with the necessary training and experience through established academic or other channels. The kinds of people needed are not ordinarily described in personnel classification systems, and in some cases descriptions cannot be made because we do not yet have enough experience to routinize the various duties and responsibilities. Today people who work in the archive organizations must usually have unique combinations of social research and computer skills, and have reasonably great tolerance for uncertainty and ambiguity regarding their personal status and the future of both themselves and their organizations.

Third, the rapid growth and change that have characterized the computer industry have had detrimental effects on the ability of archive organizations to develop with speed and efficiency. Late deliveries of equipment, the additional training that archival computer specialists must have in order to use the new computers, and the discrepancies between promised and actual computer capabilities have all caused significant differences between the plans of archival organizations and any current ability to meet those plans.

Finally, the development of archival organizations has been slowed by healthy debate among the organizations about the relative priorities of various ways of implementing the several goals and purposes. For example, a critical problem is to increase the use of archival facilities already available. On the one hand, we know that there is a large number of potential users; on the other hand, we also know that these potential users confront many problems in trying to take advantage of the services offered. In order to increase use of the services, should we concentrate on better ways of matching the specific needs of the user with the service capabilities of the archives? Or should we concentrate on training programs that will help users make local computing and other facilities more responsive to their data storage, manipulation, and analysis needs, and otherwise take better advantage of already offered archival services?

As another example, we know that the present set of archives does not store certain kinds of data that some potential users need. There is a companion problem that there is no searching facility for matching data needs with the specific, current holdings of the various data-bank organizations. Should we emphasize the development of more thorough coverage in data acquisitions? Or should we give higher priority to information systems that will improve the ability of users to determine whether any current holdings of archives match their immediate needs?

Important breakthroughs in the ability to make data collections more widely available to researchers and more useful for historical research are expected to come when the data-collection organizations themselves give greater attention to the problems and values involved. There are at least three alternative sets of experiences that can guide our approach to problems of data usability. One is the solution offered by the Inter-University Consortium for Political Research at the University of Michigan; the second is that developed by the data bank of the Center for International Studies at the Massachusetts Institute of Technology; and the third has been implemented, independently, by the Data and Program Library Service of the Social Systems Research Institute at the University of Wisconsin and the Social Science Information Center at the University of Pittsburgh.

The first alternative involves a highly specialized archive staff that operates between the data supplier and the data user, and requires the investment of considerable resources in editing and revising both data and documentation, as well as otherwise *cleaning* and preparing data and documentation for multiple use. This approach tends to give priority to data that are known to be of reasonably high quality and that have a large probability of high use rates.

The MIT group has developed a user-oriented computer system called ADMINS, that provides automated aids to the user for determining discrepancies between data and documentation or for locating inconsistencies within the data, thus permitting the user to move directly from cleaning to analysis tasks. Advantages of this system are that the user can change from a given level of cleaning to a given level of analysis at will; he is not constrained by the philosophy about cleaning and data preparation of any given archive organization; and he does not have to rely solely on the data and services available from archival organizations.

The third alternative assumes that there must be some kind of human interface between the data source and the user and that the user must have considerable control over the processing and analysis of archivally available data. It differs from the Consortium alternative in that the interface organization limits itself to finding and supplying to the user information required for cleaning data or for handling internal discrepancies and then systematically organizing such information for the benefit of other potential users. It differs from the approach of the MIT group in that it does not limit the user to a specific software system for cleaning and editing the data and documentation, or for analyzing the information, but rather provides or adapts from other sources any specific items of software that the users feel they need to use the available data resources. The major disadvantages of this organizational alternative seem to be that the user is dependent on the human skills of the people that the local data-bank organization can recruit and train, and that a given user may have to become familiar with several computer programs that perhaps involve very different procedures and assumptions; in short, the programs do not necessarily constitute a *system*.

Present information indicates that archives with high use rates give first priority to the various needs of their users and, in the process of providing service, *buffer* and otherwise protect their users from the inadequacies of the social science information system as it exists today. At this time, it seems a reasonable hypothesis that the general-purpose, local-service archives will become the major sources of data and services to users. Reasons for this apparent trend are that local users are so diverse in their needs that national organizations can respond only in gross ways to their problems, and that the computer and other facilities available to these local users require special kinds of attention to almost unique local problems.

Reference

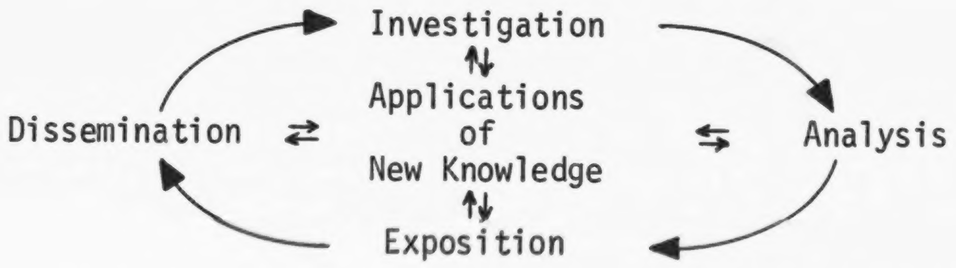
1. A description of the archive organizations that are members of the Council of Social Science Data Archives is contained in *Social Science Data Archives in the United States*, a publication of the Council.

DAVID W. CARDWELL

Oak Ridge National Laboratory
Reactor Division
Oak Ridge, Tenn.

NEW APPROACHES TO HANDLING TECHNICAL INFORMATION

The position from which I view and assess new approaches for improving communication of scientific and technical information is largely that of one who is intimately associated with the original generation of such information by means of investigation and analysis, and the practical application of it to large-scale engineering development projects. From such a vantage point, I find it most profitable to look upon each segment of information handling as one of several successive steps that should merge with one another into a closed loop wherein there is ample provision for continual interactive exchange with applications of new knowledge at its core, i.e.,



When one examines the nature of the services offered by information centers such as the 20 centers at Oak Ridge, the fundamental dilemma of technical information transfer stands out clearly. Where there is a confined range of focus, we can fairly readily machine compile, analyze, and selectively retrieve factual data. However, for broad fields of interest, our abilities are extremely hard put to proceed much beyond some systematic means that selectively guide a user to lists of publications that hopefully contain the information he seeks. It is my belief that we will never achieve much success in closing this information exchange gap until we institute measures that intelligently screen our sources of technical material to separate wheat from chaff. It is my intention to later propose a specific means for reducing in-flow to technical information centers, but since any shift in that direction will be difficult to accomplish, let us first examine some means that are being employed to cope with our present state of partial inundation.

The Automated Library primarily serves its users by compiling bibliographic reference listings, organized and stored for rapid selective search according to the nature of subject matter of publications contained therein. The most popular device for machine retrieval coding involves the use of a glossary of keywords that may be fed to the computer in various combinations to query the file. If all works well, the user obtains a selective listing of publications, including brief abstracts of their contents, from which he can proceed to requisition a library where documents are distributed either as originally published, or as some sort of photoreproduction.

The Nuclear Safety Information Center (NSIC) at Oak Ridge is an excellent high-quality example of how far it is practical to extend the Automated Library concept, when augmented by professional *guides* or evaluators. By extensive computerization, NSIC now handles 12,000 new accessions per year. Its personal services include the preparation of state-of-the-art reports, abstracts of the nuclear safety literature, indexed bibliographies of accessions, retrospective bibliographies, special reviews, and handbooks. NSIC also operates an SDI program, answers technical inquiries, and provides technical consultation. Popularity of its more personalized services has demonstrated that scientists and engineers active in the nuclear safety field need more than bibliographic listings, keyword coded, as a resource for keeping up with technical developments. The most striking response has come from evaluations made of the SDI program. This program, initiated in October 1965, now has 1700 users. Our records show that the cost of providing this highly personalized current-awareness service from a biweekly scan is about $55 per user per year. A recent survey disclosed that a substantial number of users would be willing to pay a charge within that range.

It should be obvious that to bridge the Great Divide separating our comfortable bibliographic reference referral services to workable means for gaining quick access to specific useful new knowledge *will require difficult compromises* of enforced selectivity.

About three years ago, we began the development of a computerized system capable of storage and on-line retrieval of large volumes of unstructured technical information contained in license applications for U. S. nuclear power stations. The system is known as CHORD-S, Computer Handling of Reactor Data for Safety. CHORD-S experience has shown that, by far, the greatest amount of effort and the largest number of problems arise in determining which bodies of information should be considered important enough to be included in the data bank, and how the files should be structured for logically connected search relationships. A multihierarchic structure proved satisfactory in CHORD-S as a framework for search to selectively probe different levels, but contrasting degrees of detail considered necessary by different potential users of the system required expensive and time consuming reworking of the data bank. Resolving such differences of opinion is difficult but, in retrospect, we might observe that their existence seems to allow some latitude in painful discretionary choices that lie ahead for information technology.

In spite of enormous appetites of modern automated information storage devices, we cannot build communications channels with capacities to accommodate flows that can be maintained free from sedimentary obstructions. We have little choice but to introduce positive rejection filters and diversions into the headwaters of our information streams, far more effective than any employed heretofore. Of the many options worth exploring (any of which will likely be offensive to most of the technical community), we should seek devices that induce minimum restraint on traditional rights of self expression, yet invoke screening with positive incentives to encourage high-quality reporting of useful new knowledge.

I would like to see a few pilot projects, of limited scope, having the specific objective of developing practical techniques for reducing representative narrative technical reports to provable useful facts contained therein, and ordering these facts into a pyramid-like structure for interactive computer storage and retrieval. Toward the apex of the pyramid would be concentrated the most worthwhile factual revelations from contributors, and successively lower layers would be relegated to peripheral observations that qualify, or set boundary conditions, etc., on those facts. Such a file structure can be iteratively searched to varying depths selectable by individual users of the system. One of the main departures of this notion from more traditional information system practices is the screening and ordering of subject matter toward the specific objective goals of potential users rather than attempting to digest all segments of every report published. The selective pyramid approach can provide an opportunity to condense and selectively restructure worthwhile published information without assuming the overwhelming burden of preserving every phrase of the author's prose. Information specialists who can perform such surgery will have to be knowledgeable in the technical field and be capable of facing inevitable accusations of censorship. If the technique proves successful, however, authors of technical reports might be persuaded to participate later on.

It may appear that I have ignored or neglected to recognize some very real software/hardware problems known to be associated with computer input/output manipulations. The complexity of such problems will vary considerably, depending on what we can accomplish toward screening and restructuring source material. Even so, there appear to be no overwhelming obstacles. Substantial reductions in the time required for a scientist or an engineer to locate specific information he seeks will, in my opinion, involve further application of now well understood principles of interactive computer access from remote terminals, particularly those having video-screen capability. The cathode-ray-tube terminal, as a two-way window to the internals of a computer-stored data bank, is an ideal medium for rapid trial-and-error search when followed by hard-copy retention of selective readout. The cost of equipment to provide that capability has been increasing. There are, of course, many stumbling blocks along the way of adapting telecommunications terminals to requirements for an effective information storage and retrieval system, particularly in the matter of matching human response characteristics to those of machines, and achieving component compatibility for both language and speed of transmission. However, our experience indicates that our most serious impediments to progress arise from the inability or unwillingness of individuals to extend themselves across boundaries that isolate the overspecialized disciplines of information technology, electronic communications engineering, and computer science (hardware and software).

RALPH L. DARBY

Battelle Memorial Institute
Information Operations Division
Columbus, Ohio

KNOWLEDGE-TRANSFER TOOLS OF SCIENTISTS AND TECHNOLOGISTS

I focus on the future of scientific and technical communication from the operational base of Battelle's information analysis centers. I relate three areas--scientific and technical journals, informal communication media, and critical reviews--to the acceptance of future knowledge-transfer systems and media. These three areas were chosen because they are interrelated and, more importantly, they encompass many of the communication methods of both scientists and engineers.

Battelle's information analysis centers have the advantage of operating in a research laboratory environment. They are able to continuously observe scientists and technologists at work with information and data, not only in connection with the centers' activities, but also in the laboratory. They experiment with new systems and with refinements of old ones, and obtain the reactions of real users to these systems. This broad spectrum of user audience provides many insights into the working habits of members of the scientific and technical communities.

The main responsibility of the information analysis centers is knowledge transfer. This forces the centers to be quite pragmatic in their approach to information systems. Sponsors become concerned if too large a proportion of available funding is consumed by the *information system* per se.

The tasks of scientists and technologists in the center environments include the preparation of state-of-the-art reports, technical compilations, data handbooks, and answers to technical inquiries from Government and industry. They are faced with limited budgets and, frequently, with limited time scales, particularly when *go--no-go* decisions depend on the results of their analyses. Under these conditions, frustrations, delays, excuses, or lack of information-system response are not tolerable to the technologist or scientist. All systems must be *go*!

Journal Literature

The basic functions of the journal literature, current awareness dissemination and archival storage, are well known. However, journal literature has other roles in technology transfer that are not so well understood. One of the roles that cannot be filled by either informal media or by the more advanced information systems is that of continuing education. Technologists and scientists call this *keeping up* or *keeping abreast*. Faced with the mounting tide of publication, members of the scientific and technological communities have been forced to become more selective in their reading habits.(1) Nonetheless, each scientist and technologist, in his individual way, continues to absorb and assimilate some portion of the current literature. The synergistic effect of this acquired knowledge should be studied in depth when considering possible loss of our current journals or in designing new knowledge-transfer systems. Certainly one has to agree with de Solla Price that "technology tends to accumulate more in the minds of its practitioners than it does in a well ordered literature."(2)

Many small firms and independent inventors do not have the means to tap into the large information networks, nor do they have the means for establishing their own information storage and retrieval systems. This large and diversified audience must not be ignored. Charpe, in the Fifth Annual Lecture of the Science of Science Foundation, pointed out that while over 80 percent of the research and development dollars were spent by only 200 American firms, two-thirds of the basic discoveries which resulted in important innovations came from the independent inventors and small firms.(3) Knowledge must be made available to this segment of the scientific and technical (S & T) community on a regular basis. At present, journals represent the chief means of communication with this segment of the user audience.

On one hand we are concerned about the tremendous growth of the literature and its apparent decrease in quality, while on the other hand we are concerned about the financial failure of journals in the face of economic crises. It is probably true that many journals, perhaps some of our most desirable journals, will fail. I would suggest, however, that those remaining will be of greater value and capable of more support from the S & T community. Competition for available page space will cause increased quality in those papers that are published.

Informal Communications

Both the information science community and some portions of the S & T community are studying the role of informal communications media in modern science and technology. Several notable experiments have been carried out under the auspices of the National Institutes of Health.(4, 5, 6)

While not generally recognized, an alert information analysis center capitalizes upon the informal communications media. Instead of responding to inquiries by means of bibliographies, abstracts, or indexes, a professional peer relationship can be established either through correspondence or by means of the telephone. Through these relationships, much advance information and data are exchanged. Alert scientists often detect trends and needs of the user audience at an early stage, thus assuring that the more formal products of the center, such as reports and compilations, are timely and meaningful to the user audience.

Recognizing that the scientist will use informal means of obtaining the information he desires, particularly when recent research is involved in a rapidly moving technology, the centers make such informal communications a part of the record. Trip reports, reports of telephone conversations, and correspondence become an important part of the centers' acquisition program.

There are, of course, limitations to the informal media. Total reliance cannot be placed on informal communications. The scientists or technologist must have the backup of a good information facility plus his own knowledge of the subject. But, in spite of these limitations, the advantages of informal communications are obvious. The interchange is often enjoyable and frequently leads to a rewarding interchange of ideas. Often the scientist or technologist will obtain new information or unpublished data. His correspondent will identify key papers that will enable the scientist to avoid reading bibliographies and stacks of papers in the hope that he may find the information he is seeking. Is it any wonder that informal communications are important to the scientist?

As our national networks become larger and more complex, the user audience will seek to more fully utilize his informal channels. User charges will accelerate this trend since direct contact with the user audience will be replaced by the librarian or the purchasing department. It is my belief that the continued existence of the informal channels of communication should be recognized and advantage taken of them. The information community should learn ways of tapping these transfer routes, relieving some of the strain upon the more formal transfer methods.

Critical Reviews

The *critical review* is an old but still modern approach to the scientific method. Bering of the National Institutes of Health defines a *critical review* as "a synthesis of existing knowledge relevant to some defined problem. It examines all of the hypotheses or theories germane to the question in hand, the evidence which gave rise to the theories; it assimilates new research results and other pertinent data into these theories, and identifies assumptions used to support the theories."(7)

Without the evaluation and compression of critical reviews, our scientists and technologists are faced with the full impact of the literature explosion. Critical reviews in areas of scientific and technological endeavor can do much to reduce the impact of the literature explosion.

However, it is necessary that such reviews be written by recognized experts if they are to be authoritative. The difficulties in convincing such experts to undertake comprehensive review tasks are well known. As stated by Overhage, enough critical reviews are not being written "because the incentives offered for writing them are less attractive to qualified scholars than the incentives for doing original research."(8)

A modern-day counterpart of the critical review, within the information analysis center, is the state-of-the-art report.(9) Such reports may be discipline-oriented or mission-oriented. Their proper preparation is a laborious and demanding process requiring the utmost of an individual or a team of skilled scientists or technologists. Many skilled investigators do not relish such assignments and seek to avoid them in favor of lab-bench assignments.

We have found that by offering adequate compensation, recognition, assistance through the center, and stressing the importance of the task, many scholars can be persuaded to undertake the preparation of state-of-the-art reports and critical reviews. Their first effort, while it may be approached with reluctance, can be a rewarding experience. If this is indeed the case, they will frequently seek additional opportunities for similar assignments.

Since state-of-the-art reports are frequently prepared in a rapidly evolving area, they may require revision within a year or two of their original publication. I believe this phenomenon to be of considerable importance. Time after time the centers at Battelle will receive contributions of unpublished information or data only after the importance of such information was stressed through a state-of-the-art analysis. The decision as to whether or not the report should indeed be revised and republished is dependent, of course, upon a continuing need by the user audience.

Conclusions

The ingenuity of the scientist or technologist in utilizing the information tools available to him should not be overlooked, nor should his needs and desires be overlooked. Both concentrated user audiences and diverse, decentralized user audiences must be considered.

The ultimate decisions regarding which of the journals will remain, and which media of scientific and technological communication might ultimately replace our present methods, rests not with information scientists and documentalists but with the user: the scientists and technologists whom we serve. They will accept or reject our solutions, not by direct statements, but by either a very passive acceptance or a lack of use, dependent upon how our solutions fulfill their needs. They will not be vocal in their acceptance of new services, new systems, or innovative methods. They will be vocal, at least among themselves, of their rejections. And, they will continue to develop their own paths of communication and exchange.

Careful consideration of the working habits, the methodologies, and even the peculiarities of our user audiences should be made if our developments are to be of real value to them. Above all, rigidity of system operation will hamper rather than aid the individual user.

References

1. C. Herring, "Distill or Drown: The Need for Reviews." *Physics Today*, Vol. 21, No. 9, September 1968, pp. 27-33.

2. *National Information Center*, Hearings before the Ad Hoc Subcommittee on a National Research Data Processing and Information Retrieval Center of the Committee on Education and Labor, House of Representatives, 88th Congress. Washington, D. C., Government Printing Office, 1965.

3. R. A. Charpie, "Technological Innovation and the International Economy." *Science Policy News*, Vol. 1, No. 1, July 1968, pp. 1-2, 4-6.

4. D. Green, "Death of an Experiment." *International Science and Technology*, No. 65, May 1967, pp. 82-88.

5. W. V. Thorpe, "International Statement on Information Exchange Groups." *Science*, Vol. 155, March 10, 1967, pp. 1195-1196.

6. Aries Corporation, *Interferon Scientific Memoranda - A Report on the Feasibility of Increasing the Efficiency and Effectiveness of Scientific Research through the Use of New Communications Media.* National Institute of Allergy and Infectious Diseases, National Institutes of Health, September 18, 1969. PB 184,650.

7. E. A. Bering, Jr., "Critical Reviews: The Sponsor's Point of View." *Journal of Chemical Documentation*, Vol. 8, No. 4, November 1968, pp. 236-238.

8. C. F. J. Overhage, "Science Libraries: Prospects and Problems." *Science*, Vol. 155, February 17, 1967, pp. 802-806.

9. R. L. Darby and W. H. Veazie, "Writing a State-of-the-Art Report." *Materials Research and Standards*, Vol. 8, No. 5, May 1968, pp. 28-32.

MARTIN J. DAVIDSON

North Texas State University
Denton, Texas

INFORMATION'S FATE AND FUTURE IN THE MARKETPLACE

If the concept of *total systems approach* has anything to offer an innovative and creative seminar on progress in scientific publication, it is in the suggestion that we must go far afield even to solve problems that seem straightforward. Increased sophistication in the dissemination of information and in the exponential costs associated with such sophistication may indeed not provide all the needed solutions. This is not to argue against refinements, even costly refinements in the field of providing information, but if we are to achieve truly significant breakthroughs, we must look at the larger picture.

The larger picture is that we are not dealing with scientific publication per se, but with the present and future generation of superior social, commercial, and military technologies. Our goal should be one of accelerating the generation of such technologies. Growth, narrowly defined as increased real income per capita, depends on technological progress and on those factors that promote or obstruct it.(1)

The great bulk of our usual concern regarding scientific and technical information has had to do with the supply side, the problem of *communicating-out*. I submit that we should put our focus on the demand side, not only for the purpose of determining who are the users, but for the purpose of energizing the *communicating-in* of scientific and technical information. Determining who, in fact, are the users or potential users becomes a clue to the next, more important step of getting them into action and keeping them in action.

The departmentalization, stratification, and institutionalization of industry, government, and the university mask the fact that the user of scientific and technical information is not a *consumer* at all but a type of *entrepreneur*. The distinction is between an essentially passive role and an aggressive, self-actuating role motivated by the implicit or explicit wish for either wealth or recognition.

Hooking the *communicating-in* to the *communicating-out* and profitably exploiting this interrelationship is technology transfer, a term that includes both a supply and a demand concept, which I find more meaningful than the one-sided, supply-oriented *dissemination and communication of information*. It is of course obvious that "there must be a far better climate, more receptivity to innovation and change, if there is to be more effective transfer."(2) Undeniably, the cost of pursuing technological change is extremely high both in resources and in altered social values but, as Prof. Di Gregori of the University of Houston is fond of saying, the cost of not pursuing technological change is even higher. He cites the last 200 years of Chinese history as an example.

China was the only nation in the 18th Century capable of emulating England's industrial revolution, but did not because of the high cost in terms of her values.

If we do not cease to be a pragmatic nation, we shall somehow continue to devise, however painfully, new institutions to facilitate technological innovation even at the sacrifice of existing institutions and existing social values. We must and will do whatever is necessary to energize the forces that provide technological innovation.

Resistances to specific innovations (and resistances to a social mechanism that could greatly accelerate the establishment of radical new technologies) really do not concern themselves with the matter of technological usefulness. As Barton Hacker points out, controversy (and resistance) arises in the perception of an innovation's social consequence (or the social consequences of a mechanism that could create a stream of innovations).(3) The actual utility of an innovation (or an innovation-creating mechanism) plays a decidedly secondary role. A case in point is the defeat in Congress of the Department of Commerce's 1963 Civilian Industrial Technology program. That proposal was designed to encourage and support additional research and development in industries that the Department regarded as technologically lagging. Industrial groups successfully opposed the bill because, with good reason, they feared that government sponsorship would upset existing competitive relationships.(4)

New technology that would rapidly erode asset values will also be resisted and delayed. I do not claim that this resistance need be a conscious conspiracy; it's just that laws, values, the tacit understandings of enlightened self interest, all provide the mechanism for much technological resistance. Effective resistances to technological innovation may also come from within the ranks of fundamental and applied science. A vested interest in an idea can afford resistance to change just as effectively as a vested interest in potentially jeopardized income-producing property.

The balance of this paper is an attempt to sketch the outlines of a new institution which, while tending in some ways to conflict with established values of our capitalist society, does offer a hope for an order-of-magnitude increase in our capacity to create new technologies. As a derived demand there can be expected to be a heightened application of scientific and technological information.

The Technological Risk Assumption Corporation (TRAC) is visualized, in its simplest sense, as an institution that guarantees all funds borrowed for initiating and maintaining new science-based companies. The borrowers are primarily scientists and engineers but may also include specialized marketing and financial personnel as needed. Because the TRACs must be heavily involved in making short- and long-term forecasts of all kinds (technological, market, financial, etc.), it appears advisable that they be departments or subsidiaries of established insurance companies, and staffed with expert personnel in various fields including scientists and engineers specializing in several specific areas or industries. Because of the great risks involved and as a consequence of the great potential national benefits, the ultimate guarantor underwriting losses of the TRACs must be the Federal government. TRACs would operate as a federally regulated industry with a guaranteed return on the quantity of funds that are guaranteed.

The TRAC system would obviously be costly to the government, but this cost must be measured against the social benefits that can be expected through energizing the creation of science-based companies and industries. The TRAC system is fully harmonious with a continuation of the present venture-capital financing of new industry. Private capital could also invest in the science-based companies and assume bank loans after early critical uncertainties disappeared. The TRAC concept could also be employed in the expansion of small existing science-based companies.

Though the TRAC concept might be attacked as negating capitalism, it should be noted that it facilitates the formation of large numbers of new scientific and technological entrepreneurial enterprises and stimulates the creation of new technologies. In so doing it provides a derived demand to creativity, to research and development, and to broader and more effective use of the scientific information with which this conference is concerned.

References

1. Leonard Silk, *The Research Revolution.* New York, Mc-Graw-Hill (paperback), 1963, p. 155.

2. See Charles Kimball, "Technology Transfer," in *Applied Science and Technological Progress.* National Academy of Sciences, 1967, p. 350.

3. *Technology and Culture,* April 1969, p. 198.

4. Edwin Mansfield, *The Economics of Technological Change.* New York, W.W. Norton & Co., Inc., 1968, p. 231.

RICHARD A. DERSHIMER

American Education Research Association
Washington, D. C.

ACTIVITIES TO ENHANCE COMMUNICATION IN THE AERA

General Information about AERA

The American Educational Research Association (AERA) is the largest and most representative association for the scholars and researchers concerned with educational issues and problems. Almost 70 percent of the nearly 10,000 members are academicians, the three largest subcategories being educationists, psychologists, and other social scientists.

The rapid growth from slightly more than 3,000 members in 1964 to the present number has involved concomitant changes in member characteristics. Although the percentage of representation from public schools has held constant over this period (about 12-14 percent) the generalists, such as assistant superintendents of instruction or curriculum coordinators, have been replaced by specialists in research and evaluation. There also has been an increase in the number of representatives from profit and nonprofit businesses and industries.

Evolution of the Formal Channels of Communication

In the last five years, AERA has altered several components in the three bases of its formal communications network: publications, the annual meeting, and training programs.

Since 1964, AERA has added to its publications the *American Educational Research Journal* (a scientific journal), a *Handbook of Research on Teaching*, a *Curriculum Evaluation Monograph Series*, and a seven volume series of readings. In the planning stages are an *Annual Review of Educational Research* and a quarterly publication patterned after *Psychological Bulletin*.

Thus, the Association is vigorously extending the network for data transmission, review, and assimilation into theories for both researchers and practitioners.

Since 1964, the Association's annual meeting has shifted from an aggregate of scientific papers read to report research in progress, to a meeting at which, in addition to the above function, leading scholars and researchers in education can comment upon the direction and quality of research and its relevance to the major issues facing the educational enterprise.

In the past four years, AERA has operated thirty-five 5-day training sessions on such topics as multivariate analysis, Bayesian statistics, anthropological field study methods, and evaluation of instructional materials. One year more than 750 persons were enrolled in eleven sessions. Moreover, a task force funded by a grant from the U. S. Office of Education is presently surveying needed kinds of training and designing a variety of programs appropriate to AERA support.

Studies of the Communication Networks

In an effort to understand better the effectiveness of these changes, and to postulate other needed improvements, AERA has sponsored several studies and conferences.

The most extensive study has been conducted by Johns Hopkins University Center for Research in Scientific Communication (CRSC) and has been described in two preliminary reports.(1) Dr. William Garvey and his associates have plotted a flow chart of the process by which the typical educational researcher exchanges information with his colleagues. The information exchange that takes place at the national meeting was analyzed, as were the dissemination processes which use the scientific journal article as critical medium. The resulting data were compared with information exchange activities of educational researchers in other disciplines.(2)

These studies have led to several modifications of the annual meeting, the most important being the distribution of the abstracts of papers as early as possible prior to the meeting itself. Other changes were made to facilitate the informal exchange of information.

The second study, a survey of the readership of the *Review of Educational Research,* was conducted by Linda Harris of the Institute for Communication Research at Stanford University. These data led to the conclusion that (a) an annual review of educational research was needed, and (b) educational researchers could benefit from a journal that published unsolicited reviews.

The need to understand better the flow of informal communication led to a study of invisible colleges. This investigation by David Lingwood at the Institute for Communication Research at Stanford University used sociogrammatic data to compare educational researchers to attitude researchers. While the study revealed some interesting aspects of the behavior of leaders, it did not conclusively establish that there are any invisible colleges among educational researchers.

In November 1968, AERA assembled a group of sociologists of science, communication researchers, association executives, and educational researchers to examine in what manner the characteristics of social systems in the behavioral sciences influence the direction and the quality of research. The purpose of this Colloquium was to determine what an organization such as AERA could do toward more effective methods of organization of the society of scholars in education. The papers read at the Colloquium, a compilation of interviews with leading sociologists, and a recapitulation of the discussion will appear in a final report. In progress is an evaluation of the Colloquium's impact on AERA and its affairs conducted by staff members of the Bureau of Applied Social Research, Columbia University.

Problems and Frustrations--A Summary

The American Educational Research Association has shown remarkable vigor in expanding the formal communication network to better serve its constituency. It has done so, however, in largely traditional ways: for example, by adding new publications, by allowing its regular publications to expand in size and number. (The technical journal has doubled its pages per issue in three years; and the official newsletter, the *Educational Researcher*, has increased from 49 pages in 1965 to 114 pages in 1968.)

All this has raised some serious financial questions. While it has been easy to sell almost any publication in education, continued increases in printing costs, coupled with the fact that members are no longer willing to donate their services, makes it problematical that AERA will be able to continue to expand its publication program--or even maintain what it now has.

Realizing this, the Executive Officer and the Executive Board have tried for more than two years to locate a researcher who would be willing to attempt a comprehensive study of the complete communications network. Such a study would furnish the Association with requisite information to explore installation of the latest technological developments and innovations in the storage, retrieval, and dissemination of technical information. The Executive Officer continues to seek a research unit to undertake this study.

References

1. William D. Garvey, Carnot Nelson, and Nan Lin, *A Preliminary Description of Scientific Information Exchange in Educational Research.* Baltimore, Maryland, CRSC, Johns Hopkins University (Unpublished study). ________*The Postmeeting Dissemination in Scholarly Journals of Materials Presented at the 1968 Annual Meeting of the American Educational Research Association.* Baltimore, Maryland, CRSC, Johns Hopkins University, April 1969 (Unpublished study).

2. The data were based on eight other professional societies currently participating in the CRSC program. In addition to AERA, they are American Sociological Association (ASA), Association of American Geographers (AAG), American Geophysical Union (AGU), American Meteorological Society (AMS), Optical Society of America (OSA), American

Institute of Mining, Metallurgical and Petroleum Engineers (AIME), American Institute of Aeronautics and Astronautics (AIAA) and American Society of Heating, Refrigerating, and Air-conditioning Engineers (ASHRAE).

JAMES L. DOLBY

R & D Consultants Company
Los Altos, Calif.

DEVELOPING SOFTWARE INTERFACES FOR INFORMATION PROCESSING

The role of information processing in the development of civilization is central, ancient, and obvious. If the term *information processing* is understood as a generic descriptor for the principal subprocesses of storage, transmission, and retrieval of information, then it is clear that there have been just four major advances in information processing capability in the past 6000 years. They are: 1) the invention of writing systems in the Near East in the fourth millenium B.C.; 2) the invention of alphabetic writing systems by the Greeks about 800 B.C.; 3) the invention of movable type and its application to printing about 1450 A.D.; and 4) the invention of high-speed, general-purpose digital computers about 1950 A.D.(1)

Each of these inventions provided order-of-magnitude increases in the capabilities of their users to store, transmit, and retrieve information. Each of the first three was directly followed by major changes in the nature and capabilities of the societies that exploited them. Although it is too early to assess accurately the impact of computers in this sense, historic comparisons suggest that the magnitude of the computer's effect will still be greater since it provides greater increases in information processing efficiency.

Each of the problems of information storage, transmission, and retrieval has its engineering, economic, and algorithmic coding aspects. R & D Consultants work in the latter two areas.

The problem of storage includes questions related to the ways in which natural languages code information. Since the redundancy of English is approximately 75 percent, it is evident that there are significant economies that are theoretically attainable in the storage of natural text. We have studied this problem in several forms(2, 3) and we are currently preparing a report on subject heading compression for the Library of Congress. Preliminary results indicate that elementary algorithms that take advantage of the affix-kernel structure of English(4, 5) can provide substantial word compression without any information loss.

Information stored in the form of printed text must be composed. Traditional composition techniques include certain conventional modes of representation that are primarily of an esthetic nature although there is some evidence that they also improve the reader's rate of retrieval. With the advent of high-speed composing devices, it has become necessary to develop methods for automatically performing these operations. Some of them, such as page formatting, require information inputs in addition to the text itself, and these tend to admit a relatively routine automation, although the matter of efficiency is critical and often depends primarily on the nature of the programming language used to implement the algorithms. This in turn leads to problems of computer language design for text processing problems. R & D Consultants is currently testing an experimental programming language which appears to have many promising features for such text processing problems.

Certain aspects of text composition appear to require additional information other than the text itself, but careful analysis shows that the hypothecated additional information is actually already present in the text and can be unraveled in an algorithmic way. For example, the composition problem of hyphenating and justifying text lines can be solved without recourse to dictionary look-up procedures to provide a process that is both more accurate and less expensive than human efforts. Such a system, developed by R & D personnel, is incorporated in the Mergenthaler *Linotron.*

We think that many of the problems of error detection in text can be solved by algorithmic procedures of a similar nature. Certainly an economically significant portion of the spelling errors, as well as certain classes of format perversions can be readily detected, *and in most cases* corrected by algorithmic means. Much remains to be accomplished in this area.

Another important problem of this type, and one that we think is accessible to existing technology, concerns composition of formulae from typescript (OCR) or typewriter-generated machinable input. Apart from its direct economic advantages, a solution would substantially decrease the considerable delay between submission of a scientific manuscript and its appearance in a permanently available form.

Problems of information retrieval have rightly commanded considerable attention in recent years. The exponential growth of information archives (documented, for instance, in Refs.6, 7) creates irresistible pressures against static access techniques. We have been particularly concerned with problems of measuring and increasing access to archival collections, and to this end have undertaken studies of computerized library catalogs(6) the structure of bibliographies(6), and the automatic creation of indexes. The latter is in many respects the most exciting and promising of the access-expanding techniques that we have investigated. Our preliminary studies have shown that human-produced, back-of-the-book type indexes display a consistent structural pattern. The distribution of text references associated with the index terms is essentially exponential(8), and the grammatical structure of the index entries themselves shows regularities which are of importance when indexing algorithms are considered.(9) Based on a number of studies conducted during the past several years, an elementary indexing algorithm has been designed, and applied to the text of a book.(6) The index that appears in Reference 6 has received minor human post editing, as is therein described. In all of its measurable characteristics, this experimental index is consistent with human-produced indexes.

Very little information has been published about the statistical and structural characteristics of indexes, although measurement of the performance of an algorithmic indexing procedure must evidently be referred to some standard. In order to remedy this omission, we have begun to collect and analyze the indexes occurring in a random sample of some 1800 monographs drawn from a university library of medium size. Upon completion of this project it will be possible to assign a degree of certainty to statements about the nature of indexes, and also to provide meaningful comparisons of performance of various indexing procedures.

It is important to recognize that an index to a text is a global function of the text. This means that it depends on the text considered as a whole, rather than on consideration of consecutive local portions of it with simple accumulation of the locally determined index terms. From this point of view, indexing algorithms contrast sharply with editing or hyphenation or page composition algorithms which rely on relatively small amounts of context in order to make their decisions. One consequence of this difference is that indexes based on literature occurring in a specified field of interest can be cumulatively recompiled (from a greatly compressed subset of the original texts) to produce a cumulative index which is not simply the union of the indexes of the papers and monographs that constitute the literature base. Currently, such cumulative indexes are not published. However, with the availability of high performance and low cost automatic indexing techniques, it will no doubt become desirable to construct and publish such cumulations on a regular basis for literature in important and rapidly changing fields.

There is another class of retrieval problems that demands attention. In order to be effective in typical environments, a retrieval system should admit incorrect input requests; that is, it should have some degree of stability with respect to the knowledge of the system's user. An elementary but important example of the need for stability is provided by incorrectly spelled author's names given as retrieval inputs; another relates to spelling errors in author or title designations that are due to mechanical errors rather than wrong or incomplete information. Indeed, it is not too much to hope that error correction techniques of a linguistic nature could be included in computer operating systems to assist in program debugging in a positive manner, but this will require a redundancy increase in the format of most higher level software systems.

References

1. J. L. Dolby, V. J. Forsyth, and H. L. Resnikoff, *The Cost of Maintaining and Updating Library Card Catalogs*, Final Report under Contract OEC-9-8-00292-0107, R & D Consultants Co., May 27, 1969.

2. J. L. Dolby, "An Algorithm for Noisy Matches in Catalog Searching." In Cunningham, Jay L., et al., *A Study of the Organization and Search*

of Bibliographic Holdings Records in On-Line Computer Systems: Phase I, Final Report Prepared by the Institute for Library Research of the University of California at Berkeley, March 1969.

3. J. L. Dolby, H. L. Resnikoff, and J. W. Tukey, "A Ruly Code for Serials." *Proceedings of the American Society for Information Science,* Vol. 6, 1969, pp. 113-124.

4. H. L. Resnikoff and J. L. Dolby, "The Nature of Affixing in Written English I, II." *Mechanical Translation,* Vol. 8, 1965, pp. 84-89, and *Mechanical Translation,* Vol. 9, 1966, pp. 23-33.

5. J. L. Dolby and H. L. Resnikoff, "Statistics of Language: Structure of Written English Words." *Encyclopaedia of Linguistics, Information and Control.* Oxford, Pergamon Press, 1969, pp. 581-584.

6. J. L. Dolby, V. J. Forsyth, and H. L. Resnikoff, *Computerized Library Catalogs: Their Growth, Cost, and Utility.* Cambridge, The M.I.T. Press, 1969.

7. D. J. De Solla Price, *Science Since Babylon.* New Haven, Yale University Press, 1962.

8. H. L. Resnikoff and J. L. Dolby, "On Indexing Books." R & D Consultants Co. working paper, August 1969.

9. J. L. Dolby, "The Structure of Indexing: the Distribution of Structure-Word-Free Back-of-the-Book Entries." *Proceedings of the American Society for Information Science,* Vol. 5, 1968, pp. 65-72.

JOHN S. GILMORE

University of Denver
Denver Research Institute
Denver, Colo.

THE USER OF TECHNOLOGICAL INFORMATION, TARGET OR PARTICIPANT?

The Denver Research Institute's Project for the Analysis of Technology Transfer has been studying the means by which scientists and engineers acquire information on new technology from outside their organizations, and the implications of their behavior for technology disseminators.(1)

The investigation has concentrated on those knowledgeable senior people, often found in research and engineering groups, who serve as mediators between their group, its information needs, and the various external sources of information. In certain industrial firms, we found mediators in three functionally different categories: in research-oriented, product-oriented, and technical management groups. The research-oriented mediators were usually concerned with natural phenomena and analytic problems. They usually sought information about basic scientific knowledge, analytic techniques, or new products, materials, or components. Product-oriented mediators principally dealt with the application of currently or soon-to-be-available materials and components for the development, design, and production of a product. They often sought information on design concepts, technological performance, new applications, and new products, materials, or components. Technical managers were not primarily occupied in technical work. They typically needed information relating technological capabilities to market opportunities, or on the economic implications of substituting materials or components.

Charts I, II, and III summarize the rank ordering by respondents to 17 external channels of information for two types of information use, current awareness and problem-solving activities.

The respondents were drawn from four manufacturing industries with little defense/space orientation (i.e., batteries, medical electronics, industrial controls, and printing and reproduction machinery) and one technically oriented service (vocational technical education). We were interested to see the importance attributed by people in these firms to textbooks and handbooks for problem solving and to trade publications for maintaining awareness. The charts suggest the crucial role of journals for research-oriented personnel, and the importance of supplier channels (both supplier personnel and supplier catalogs) for product-oriented and technical management people.

Ready access and familiarity with a given channel correlated directly with the importance attributed to it. For example, the relatively new Federal government channels for reporting government-supported research were rated comparatively low.

CHART I. IMPORTANCE RANKING OF EXTERNAL INFORMATION CHANNELS BY RESEARCH-ORIENTED PERSONNEL

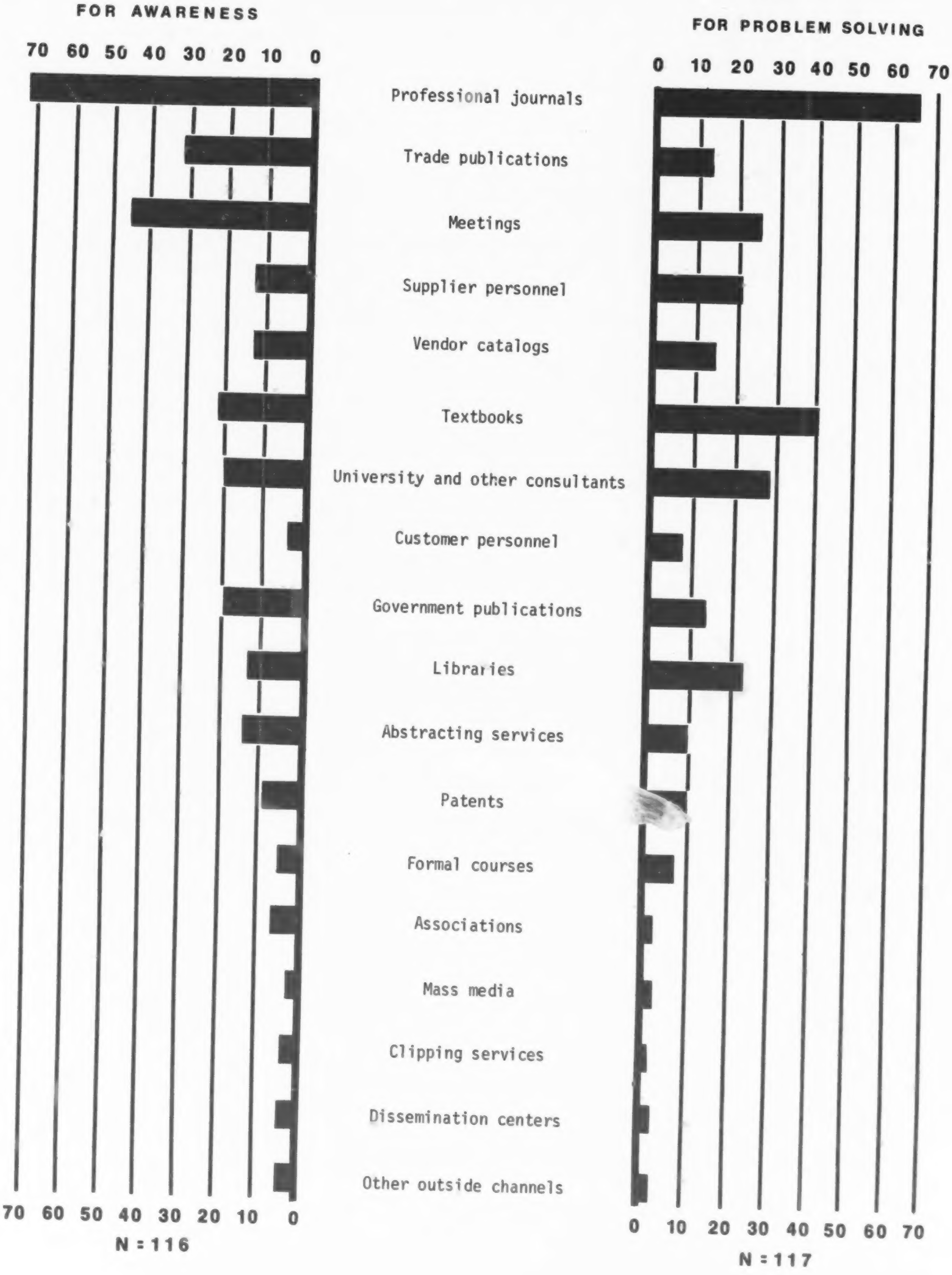


CHART II. IMPORTANCE RANKING OF EXTERNAL INFORMATION CHANNELS BY PRODUCT-ORIENTED PERSONNEL

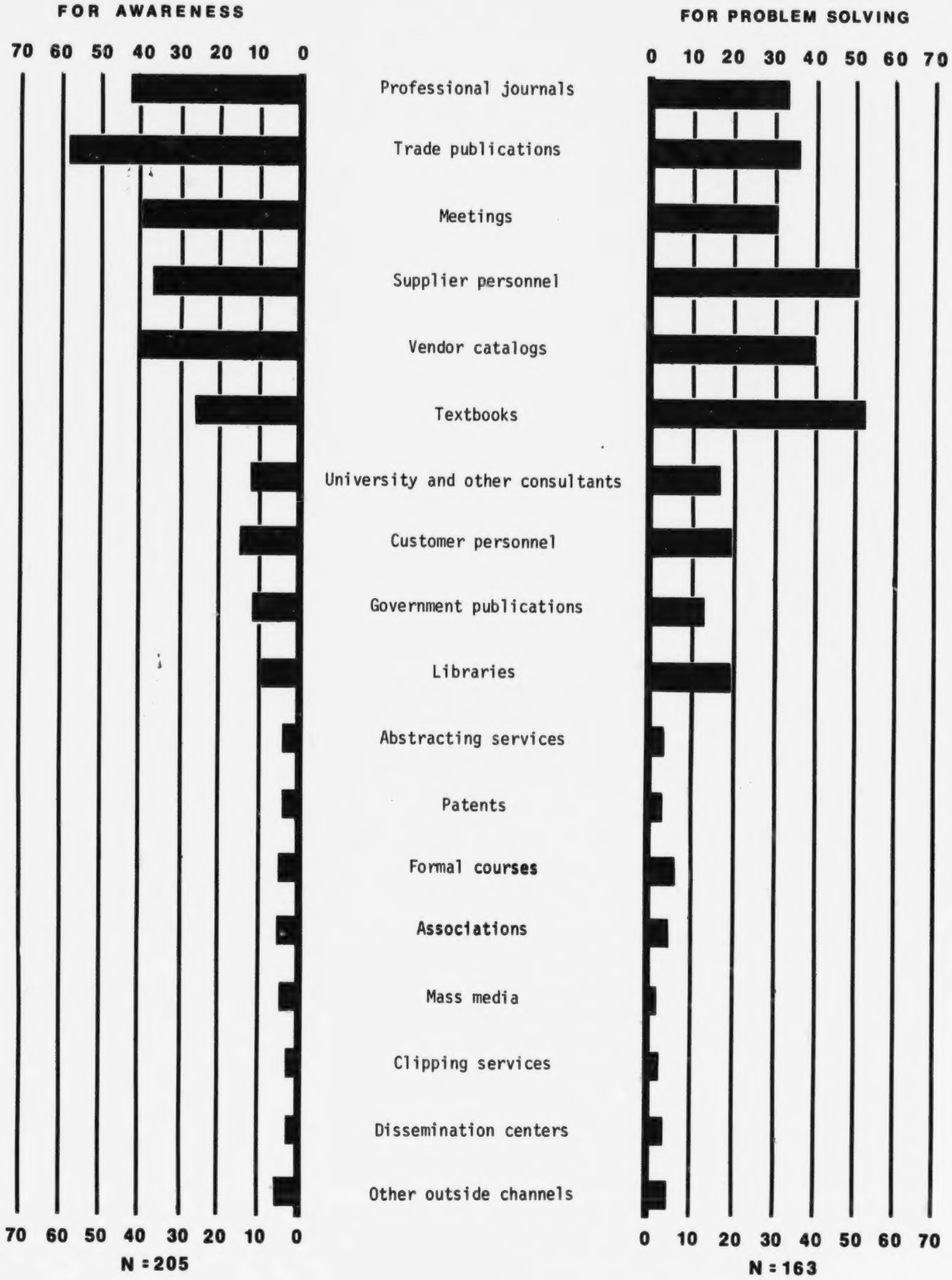
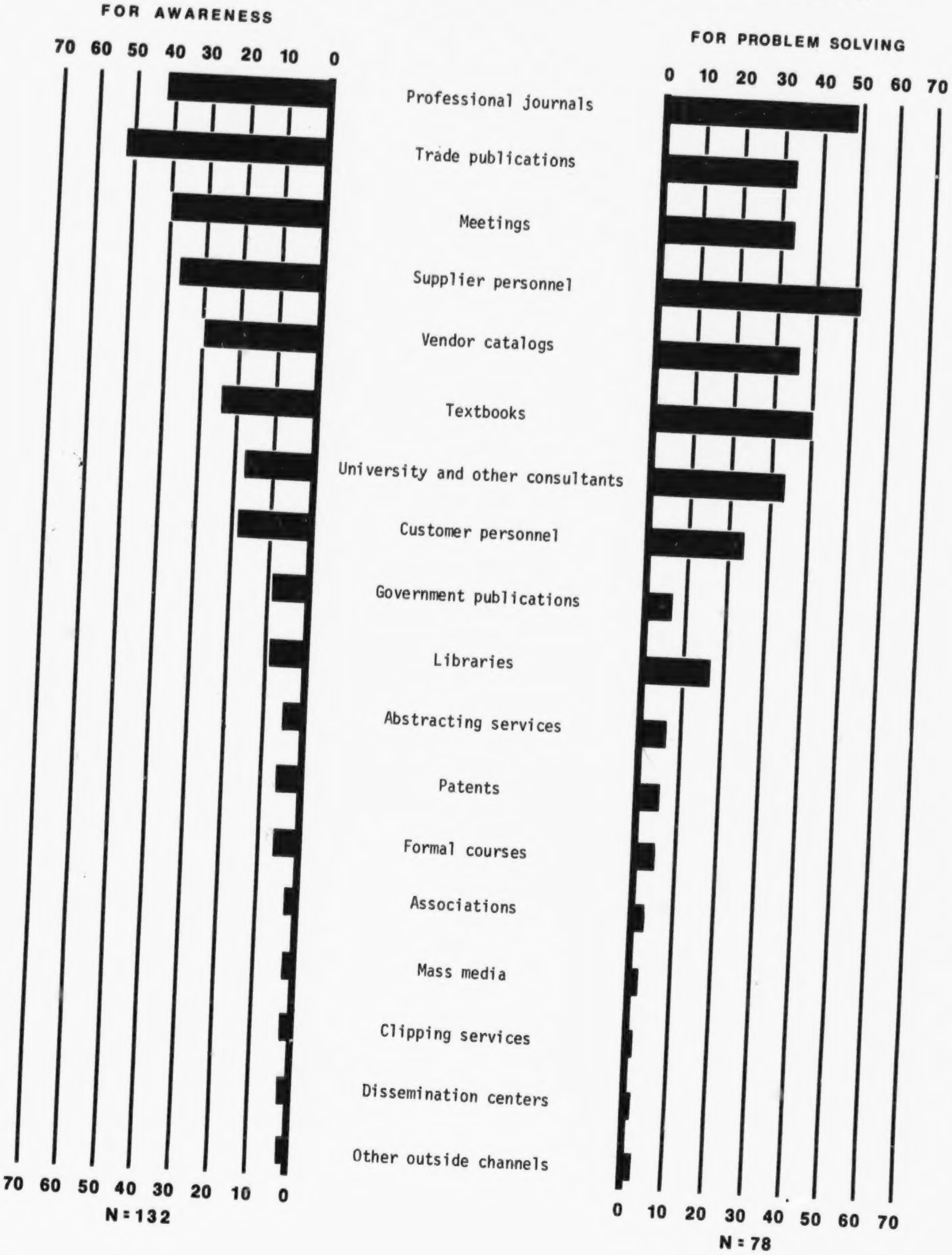


CHART III. IMPORTANCE RANKING OF EXTERNAL INFORMATION CHANNELS BY TECHNICAL MANAGEMENT PERSONNEL



FOR AWARENESS

FOR PROBLEM SOLVING

Each of the mediators tends to tap many and varied channels. Some channels complement each other and add to each other's value. Some redundancy appears desirable; it is probably important for effective dissemination, particularly if dissemination is directed at more than one functional type of R & D mediator.

Our study has led us to raise the following question: Should there be major concern with the technology dissemination function, or should the term *dissemination* be de-emphasized in favor of the term *communication* and the concepts the latter fosters? Communication requires the participation of the user. Those responsible for improving technological communication may need to consider ever more active intervention with the user or with the user's environment. The success of technological communication systems depends at least as much on the users as on the transmitters of information. Therefore, it seems logical that those seeking means for progress in this field should seek possible changes in the user situation, as well as proposing possible changes in publication and dissemination.

Reference

1. John S. Gilmore, William S. Gould, Theodore D. Browne, Carl von E. Bickert, Dean C. Coddington, J. Gordon Milliken, and John G. Welles. *The Channels of Technology Acquisition in Commercial Firms, and the NASA Dissemination Program.* National Aeronautics and Space Administration, 1967. NASA Report CR-790.

JOHN R. HART

North Carolina State University
School of Engineering
Industrial Extension Service
Raleigh, N. Car.

INFORMATION SERVICES FOR INDUSTRY

The fast changing and burgeoning character of the industry of our State is creating great technological and managerial needs. It has long been recognized that the School of Engineering has technical capabilities which, when properly channeled, can provide a vital assist in meeting those needs. We are now doing just that through the field work of the Industrial Extension Service and through the offering of timely short courses, conferences, and workshops using School faculty and appropriate industry personnel.

At present, the Industrial Extension Service (IES) and our State Technical Services (STS) Program have a combined staff, consisting of 23 professional personnel representing eight engineering disciplines. The Field Services section provides for industry liaison, referral services, problem-solving, technical assistance to small industry, and conducts most of the in-plant training courses. The extension faculty are responsible for the development and much of the presentation of the educational programs. Both of these groups are located on the campus of North Carolina State University. Participation in the STS Program by other North Carolina institutions includes the computerized literature searching services of the North Carolina Board of Science and Technology and educational programs by Wake Forest University in Winston-Salem, East Carolina University at Greenville, and the Research Triangle Institute at Durham.

The IES provides the following types of information services:

LIBRARY SEARCHES - identifying and making available book and article references on particular subjects or problems.

DIRECTORIES - providing readily available information on individuals and organizations rendering industry services as well as the location and availability of North Carolina produced parts and materials.

BULLETINS AND PAMPHLETS - containing information easy to understand and apply on numerous subjects of interest to small industries.

HANDBOOK LIBRARY - containing a wide range of handbooks in technical and managerial areas that are on loan free.

REPRINTS - including key comprehensive articles of general interest and detailed articles of specific technical interest.

IES NEWSLETTER - providing current information monthly on the extension activities of North Carolina State University as well as other articles of interest to North Carolina industry.

INFORMATION SERVICES - The information services program of the Industrial Extension Service seeks to make available to industry valuable sources and aids found in the technical literature. This is accomplished through the library and various kinds of publications.

Various programs designed to fill the continuing education needs of industry are offered through personal instruction, television and telephone media, programmed instruction materials, and a film library.

SHORT COURSES AND CONFERENCES - Short courses and conferences are conducted on campus and around the State for updating and training of industrial personnel. Courses are given in subjects as diverse as work simplification, concrete technology, construction management, fluidics, and critical path scheduling. New courses can be developed on request.

EDUCATIONAL TELEVISION COURSES - Educational Television courses are offered during late afternoon hours via the State's five-channel open-circuit educational television network.

IN-PLANT TRAINING COURSES AND MATERIALS - In-plant training courses and materials are available on supervisory and technical topics. Courses are conducted by extension staff members on supervisory, managerial and technical topics.

TELEPHONE-TRANSMITTED COURSES - Telephone-transmitted courses via the *Electrowriter* system are available to provide in-plant instruction without the necessity of frequent travel for either the student or the instructor.

FILM LIBRARY - The film library contains over 150 training films on the general topics of supervision and human relations, industrial economy, industrial engineering, industrial safety, and methods improvement.

LIBRARY OF PROGRAMMED INSTRUCTION MATERIAL (LOPIM) - Over 100 programmed instruction (self-teaching) courses are available on such topics as accounting, computers, electronics, quality and cost control, machine control, management, and statistics.

CORRESPONDENCE AND EVENING COURSES - Correspondence and evening courses for college credit are offered each semester.

Engineering Assistance

A staff of field representatives and extension specialists is available to help small industrial firms lacking engineering specialists to identify problems and to assist in solving them until the firms or some other source can reasonably complete the task.

Also emphasized is the location and application of new technological ideas, and the compilation and presentation of technical data on materials, products and processes which present opportunities for enterprise expansion or diversification.

Sometimes assistance is provided by a combination of instruction and demonstration. For certain common areas of need, an Industrial Extension Service staff member can be assigned to work with members of a firm's staff to provide *on-the-job training* based on the firm's operations and problems. For this, the industry *students* are supplied with available text and training materials.

In cases where the Industrial Extension Service is not able to provide complete assistance, it serves as a liaison between the requesting firm and faculty members, professional consultants, state and federal agencies, and other sources where the needed assistance can be obtained.

Representative areas in which assistance has been provided are:

- Feasibility studies for the manufacture of new products
- Plant layout and equipment requirements studies
- Liaison service in locating professional and technical expertise

- Manufacturing process trouble-shooting
- Verification of product cost estimating procedures
- Location of suppliers of equipment and services
- Development of quality control, inventory control, and production control procedures
- Methods improvement analysis

JACK HELLER

State University of New York at Stony Brook
Department of Applied Analysis
Stony Brook, Long Island, N. Y.

MUSEUM DATA AS AN INFORMATION RESOURCE

As people learn to make use of the information stores in libraries, archives, government collections, museums, and other repositories, today's needs for computerization and for a better understanding of how to take advantage of computerized data bases can only become magnified. The Museum Computer Network research project described here is one effort to explore the use of new tools and techniques that will make it possible for these repositories to cope with predictable demands.

Fifteen museums in New York City and the National Gallery of Art formed a consortium in 1967 to examine the feasibility of a computerized catalog of their own and other museums' combined collections, and to implement a trial data bank.

Written records of museum holdings are of two kinds, although the distinction is blurred in practice. Accession records consist of an object's identification number, data concerning its source, method of acquisition, physical description, and the like. Since no research or special knowledge goes into their making, accession records exist, at least in theory, for all museum objects. Catalog records contain data verified by professional research as to authenticity, provenance, history, and bibliography. Only a small portion of museum holdings are fully and finally cataloged.

Museum records are anything but standard. They vary in types and amount of information, how it is subdivided and labeled, the degree and method of indexing, vocabulary used, and format. Every kind of variation may be found, not only between museums, but within the individual department and the individual file. Most variations are random, but many embody policy, tradition, or opinion, and so must be kept. Underlying this diversity is a remarkable uniformity of factual content. The possibility of an organized data bank derives from the presence of comparable data, implicit or explicit, in nearly all records. The extraction of these data requires logical analysis, much of which can eventually be programmed but all of which is now done by people.

Further, most museums own materials outside their field of interest, and these orphan collections are hardest of all to find and least likely to be cataloged. A general data bank could bring them into meaningful context and lead to advantageous exchanges, sometimes of objects and sometimes of personnel able to catalog them.

The files generated to date employ programs that were conceived for organizing, storing, and retrieving data of variable length.

The programs have the general name, GRIPHOS (General Retrieval of Information Processing for Humanities Oriented Studies).

The chief file is the item file that consists of a random sequence of *items* of any length, each describing one entity, i.e., an object, person, document, or event. Items in turn are divided into any number of *units* each of which conveys one logical fact or a cluster of related facts. The unit consists of an *annotator* (a code number), showing its significance, associated with a word, phrase, number, or paragraph of factual content. The unit may be considered a statement about the subject of the item. The units are of two kinds. The shorter, stating one fact only, and likely to be repeated in more than one item, are called *descriptors*. These not only convey information, but also serve as an index to the file. They alone are the subject of search and sort operations.

The other units, known as *free text*, form a reservoir of supplementary information. They are generally longer and often unique to a single item. This, together with the conjunction of several facts in a single unit, makes free text unsuitable for routine searching and sorting. It would have been easy to combine all the free text in any one item into a single abstract, but users would too often be burdened with excess information. The text is therefore divided into as many units as necessary, each of which is assigned a code number consecutive with that of a related descriptor. For example, code #2 identifies the name of an object's owner (a descriptor) while #3 tells about the acquisition of the object by that owner (free text).

Perhaps the most important, if quite imprecise, datum we have come up with is that it costs, all told and on the average, about $1.50 to see an art museum's object record through, from card files to electronic storage. This, assuming access to some nonprofit computing facility, is probably a maximum figure. Individual items vary in length from almost nothing to hundreds of lines, but the average for records of this type seems to be 20 units and roughly 400 characters.

Some problems might arise from museums' and individuals' rights to information. Where files represent a life work, the proprietor will hardly care to see their content rushed into print by a user of a data bank. There is also simple pride; museum publications set high standards of style and accuracy. The labor needed to maintain files helps to explain why relatively little goes to press. But if making data public through the Network is regarded as a form of publication, the same obstacle will arise, and the mass of semi-cataloged material will not be any more accessible than it is today.

Some formula is needed to assure free circulation at least of basic inventory information (accession records) without researchers being deprived of credit or profit, and without institutions feeling their reputations threatened by every dubious descriptor. The question of organization beyond our experimental phase remains open.

It should be noted that a network such as this lacks natural limits, geographic, linguistic, or disciplinary. The problems of museum record keeping are those of record keeping in general, and the solutions now taking shape should apply as well to any subject matter, indeed to any mixture of subjects. Machine storage capacity is growing faster than data can be gathered. There is literally no stopping place in the whole universe of recordable information.

Bibliography

1. David Vance, "A Data Bank of Museum Holdings." *ICRH Newsletter*, Vol. 4, No. 7, March 1969, pp. 3-7.

2. Jack Heller, "The Value of a Computerized Data Bank as an Adjunct to a Museum Card Catalog," in *Computers and Their Potential Applications in Museums*. New York, Arno Press, 1968.

3. E. Ellin, *The Museum Computer Network*. 27 W. 53rd Street, New York City.

CONYERS HERRING

Bell Telephone Laboratories
Murray Hill, N. J.

CURRENT ECONOMICS OF PRIMARY JOURNALS IN BASIC SCIENCE AND TECHNOLOGY

I present here a general philosophy for approaching problems of policy for the scientific and technical literature; data from a brief survey of the economic situation and problems of primary journals; and an attempt to construct arguments for or against certain policies by combining the philosophy with the data.

Philosophy

At the outset we must recognize that the value of primary journals is measured by what they contribute to the progress of science and technology as a whole. Thus to make recommendations about their economics one should ask two interrelated questions: 1) If the nation (or the world) devotes a given amount of money and manpower to a certain broad area of science or technology, what proportion of these resources should go into primary journal publication in order to achieve the most progress? 2) Given this proportion, what policies controllable by the publisher of journals, and what policies of government agencies and other sponsors of R & D toward these publishers, will yield the most ultimate progress? In all cases, the yardstick is the health and progress of the entire research or development effort, not the mere provision of a *publication outlet* for a given number of papers. This is our first basic principle.

Who is to be the judge of what contributes, and how much, to this *health and progress*? The various entities that now make decisions about information transfer--individuals, university departments and libraries, companies doing research and development, scientific societies, private publishers, government agencies, etc.--base these decisions on perspectives of varying breadth and varying range in time, and on motives that involve altruism and self-interest in varying proportions. We believe that the diversity of viewpoints, motivations, and above all the expertise among these entities is a resource that can, if intelligently used, contribute more to the health and usefulness of primary publications than could an extreme centralization of decision making. Our second basic principle is therefore that those entities that combine a considerable economic or administrative power with some of the broadest and most long-range points of view--OST, NSF, many scientific and technical societies, etc.--should set policies that encourage innovation and freedom of choice by individuals and smaller groups, and constrain these activities only to the extent necessary to ensure that the general direction of their choices is not antisocial.

Several rather general corollaries can be drawn from these two principles. The first of these has to do with allowing the users of information to be the principal judges of its value to them, as compared with other things that society could provide them at the same cost. So far as is feasible in our present free-market society, whenever any individuals or small groups are willing to pay a cost at least as great as the incremental cost of providing them with an information service (such as a subscription to a journal), they should be able to obtain this service.

Two more corollaries have to do with the economic importance of items that do not appear in the financial records of journals, but that still represent costs to society for the complete process of getting information from its producers to its users. One such item is the time spent by users. We must always keep in mind the value of the time spent by the users of journals in finding material in them and in reading and digesting it, and of the time that potential users may waste through failing to make contact with information that could help them. Another item that should not be forgotten is the *hidden cost* in journal production, the value of the time spent by referees and editors who are unpaid or inadequately paid, and of the services effectively donated by the employers of these referees and editors. No comparison of alternatives for journal production is valid without consideration of these hidden costs.

Data

Our brief survey assembled data on size, price, birth and mortality, etc. (the things identifiable by examination of issues in a library) for about 250 U. S. journals in all fields (preponderantly the physical sciences and engineering), and a considerably smaller number of foreign journals. Data on circulation and on production costs and income were obtained for about 45 U. S. journals by correspondence with the publishers.

The diversity of the journal population is overwhelming. In bulk, the largest journals publish about 500 times as much per year as the smallest. The average journal has been increasing in bulk at about 7 percent per year. The spread in price is also enormous, the *best buy* in 1968 having provided 90 times as many words per penny as the *worst buy* in our sample. A very interesting fact is that, whereas journals without page charges have usually been getting more expensive per word over the last decade, those with page charges have usually been getting cheaper. Circulations also vary widely, though few journals lie outside the range 1500 to 15,000.

New journals are being born at a considerable rate and, in the U. S. at least, the mortality of journals of the type considered here is negligible. As a result, the number of such U. S. journals is increasing at something like 6 percent a year.

Few fields are so specialized that a worker can rely on a single journal or even half a dozen for as much as two-thirds or three-fourths of his useful information. The rate of obsolescence of material published in primary journals is slower than is sometimes supposed.

Arguments and Admonitions

An oversimplified application of the philosophy outlined earlier would be to say that the support of journals should be such that they can be sold at prices that will maximize the net benefit to society from decisions of enlightened buyers as to whether or not they should subscribe to the journals. In the real world, we desire, as a second goal, that the support policy be such as to encourage correct decisions on the whole range of factors having to do with the birth, death, and quality of primary journals. Finally, as a third goal, we desire that the journals have reasonable economic stability in the face of temporary fluctuations in conditions.

A detailed analysis of each of the various combinations of pressures and decisions shows that for many of the combinations, socially beneficial decisions are encouraged by having a page charge or other subsidy of pre-run costs, while for other combinations, there can be an advantage to having a major part of pre-run costs covered by subscription income. Thus, the optimum support policy for encouraging good management is a little uncertain.

Since a large majority of the papers published in practically all fields of science and technology are supported by the federal government or by industry, we draw the conclusion that the page-charge system be widely adopted for journals published by nonprofit publishers. All government contracts for research or development should budget reasonable amounts for possible publication expenses, and these amounts should not be transferrable to other uses. If the necessary administrative machinery can be set up, page charges should be paid directly to the journals by the sponsoring agencies, without passing through the hands of the contracting institutions. The guidelines enunciated by the Federal Council for Science and Technology in 1961(1) for such payments seem to be basically sound, but will require some minor modifications in the direction of widening the scope of possible policy decisions by journals.

Besides this recommendation to the government, there are several suggestions that can be made to publishers of journals, particularly societies. One such is for federation of a number of small publishing operations into a single publishing organization that can publish a number of journals in such a way as to make possible employment of skilled personnel and maintenance of uniform work loads. Another is for offering journals at a certain price per volume, with an unspecified number of volumes per year (private publishers often do this). Another is for wider use of typewriter composition and photo-offset, or other technologies based on composition within the production office. Finally, there is the desirability of systems of accounting that clearly separate pre-run, run-off, and other elements of cost.

Reference

1. *Scientific Information Notes*, Vol. 3, No. 5, 1961, p. 1.

Bibliography

Previous studies of journal economics include the following:

1. T. H. Campbell and J. Edminsten, *Characteristics of Scientific Journals - 1962.* Washington, D. C., Herner and Co., 1963.

2. L. J. Paige, W. T. Martin, and A. Rosenberg, *A Special Report on the Means of Financing Mathematical Journals.* Providence, Rhode Island, American Mathematical Society, 1963.

3. Case Institute of Technology, Operations Research Group. *An Operational Research Study of the Dissemination and Use of Recorded Scientific Information. A Study under NSF G-8434.* Cleveland, Ohio, Case Institute of Technology, 1960. PB 171 503.

4. I. E. Stewart, "Economics of Journals." *Federation Proceedings*, Vol. 22, 1963, pp. 1002-1007.

5. *Some Characteristics of Primary Periodicals in the Domain of the Physical Sciences.* Paris, France, Abstracting Board of the International Council of Scientific Unions, 1966. *Some Characteristics of Primary Periodicals in the Domain of the Chemical Sciences.* Paris, France, Abstracting Board of the International Council of Scientific Unions, 1966.

6. *Engineering Societies and Their Literature Problems.* New York, Engineers Joint Council, 1967. A paper especially concerned with journal economics in this EJC publication is: S. W. Herwald, "Society Publications: Their Finance Problems."

7. C. Andrade, *A Report on the Page Charge Practices of 51 Primary Mathematics Journals.* Providence, Rhode Island, American Mathematical Society, 1968.

8. *Scientific Journal Page Charge Practice.* Washington, D. C., Biological Sciences Communication Project of the George Washington University, 1968.

RICHARD L. KENYON

American Chemical Society
Washington, D. C.

COMMUNICATING CHEMICAL INFORMATION

The American Chemical Society has long been a producer of primary journals and of abstracts and indexes as well as an organizer of meetings. During the past several years a growing amount of effort has been devoted to organizing and disseminating chemical information in ways that may build a base for optimum handling and use of the rapidly growing volume of material.

Study is underway on the nature of information flow to, through, and from meetings with the object of learning about the structure of actual information exchange in formally organized programs which are accompanied by a considerable amount of informal exchange and which yield irregular levels of formal transmission through publication.

The journal system has been developed with publications of varying breadth and depth. A weekly news magazine, *Chemical & Engineering News,* designed to inform accurately and lead to sources for details, cuts across the entire field of work important to chemically trained people, selecting on a news basis the most significant information. At the other end of the spectrum are research journals publishing only research papers and communications which have been critically reviewed (refereed) and edited. These journals perform alerting and awareness service as well as the archival function. In addition they form an important part of the base of social-professional structure of the scientific community, something which cannot be ignored in the organization and dissemination of information.

Another set of publications is the review journals, including both comprehensive critical reviews of selected areas of sub-disciplines and concise reviews of closely limited fields that are of high current interest. A publication of the latter type, *Accounts of Chemical Research,* contains reviews of four to six pages written by an active scientist of high authoritative standing in the field. The author is asked to write almost entirely from his own current, live knowledge and give only the most important references. The object is to present a fresh and highly current view; the journal has been well received.

Primary Publication Forms

Studies are underway on the more effective design of journals and other forms of printed information media. Possible designs include:

1. The *Short Paper* Journal--One form of this journal requires submission of both a manuscript in short form (not more than twice the

length of a *Communication to the Editor*) as well as a fully detailed form including reproducible experimental information. The latter is used in critical review or refereeing for acceptance and is filed in archives while the short form is published. The second form of short paper journals requires only submission of the short paper for publication. Additional details must be secured from the author.

2. Limited Scope Selective Journal--These journals would be developed by packaging selections of accepted papers according to group interest profiles rather than according to sub-disciplines, the basis of most chemical journals today. Expected size of user groups would be between 500 and 2500. The subject matter might often be interdisciplinary or intersubdisciplinary with overlap among the many journals produced.

3. Selective Dissemination of Individual Papers--A continuous flow of papers related to the individual scientist's interests is highly attractive in principle. If done by matching content of papers to individual profiles the content value should be high. The task of maintaining and servicing profiles for more than 100,000 scientists is presently prohibitive, but may not remain so with developing technology.

4. Title Lists with Individual Paper Availability--Circulation of tables of contents would allow users to make their own selections, but delay and processing appear cumbersome.

5. Abstract Journals with Papers Available on Order--This form offers better current awareness information than item 4, but similar disadvantages.

6. Manuscript Preprints--Manuscripts entered into consideration for publication in journals are made immediately available on order based on prompt publication of summaries. For proper functioning, this requires the authors' permission as well as treatment of material as private communications until edited and published.

Computer-Assisted Photocomposition

Four journals now are being produced by encoding manuscripts on tape for computer storage and handling. The tape is used to operate a photocomposition machine (Photon) which produces the film base for making a negative for offset printing. This system is being studied relative to its potential for use in a fully computer-based information system in which a manuscript will require but a single intellectual analysis and will be encoded only once for storage and use in producing the primary journal, abstract, index, and other forms. Editorial and keyboarding operations are being combined into a single flow to produce a machine-readable record with little variation from one publication to another.

Chemical Abstracts Service (CAS)

The world's chemical literature has been abstracted by the American Chemical Society in its *Chemical Abstracts* since 1907. Since 1962 the computer has been in use by CAS and work is underway to put the entire system on a computer base.

The CAS system is an integrated man-machine organization that takes advantage of computer capabilities whenever they provide economic or technical benefits to help attain over-all ACS/CAS objectives. This system is designed to permit easy adaptation of new hardware and new data processing techniques as rapidly as CAS experimentation and community developments provide the required technology. The following design principles are being employed:

- The system and its subsystems are designed to be independent of specific items of equipment wherever possible.
- System-wide standards are fully defined and maintained. Each data element to be processed in the system is defined and given permanent, unique identification. Program subroutines which handle these data elements are carefully documented and stored in the computer for routine use. Standard file formats identify each information record by code, define the location of the data, and state the length of each variable-length string.
- The processing workflow is modularized. Individual modules are small and simple enough to permit ready adjustment or replacement, and changes need be made only on the affected modules.
- The system is founded on a unified, carefully integrated data base which is built through a single analysis step for each document covered in the system. This data base provides information for all CAS publications and services and all searches of CAS computer files. Therefore, the data from each document are recorded in the maximum technical detail required in any CAS service or in any system based on a search of CAS files. The records reflect adopted standards for character sets, coding techniques, abbreviations, etc. For each output application, the data will be formatted to make the information content related to the specific output purpose easily identifiable by the recipient of the output.

CAS processing activities will become increasingly automated throughout the shift to a computer base, but it is essential that regular CAS services not be interrupted during this transition. To meet these objectives, a series of modifications must be made to the processing stream such that each change comes closer to the ultimate intended system while assuring continuity of the established regular services.

The CAS development program is based in seven areas of activity:

1. Computer-based, subject-oriented sorting routines to organize output data into forms easy for humans to utilize. Routines are now being developed for sorting author names and associated document titles, and for ordering subject-oriented nonstructural concepts in combination with chemical nomenclature. The first operational system is scheduled for 1970.

2. Direct computer control of every form of output display, including composition of all CAS publications. The first version of a fully computer-based system capable of handling the complete range of CAS publications is scheduled for completion by the end of 1971 with the processing of *CA Issues* through the CAS computer composition system. The system is being developed so that relatively simple adjustments of computer processes will allow the use of new composition equipment as the state of the art develops.

3. Use of computer support to simplify the keyboarding, proofing, and correction cycle and to eliminate multiple manual handlings of the same information.

4. Shift of an appreciable portion of the intellectual editing task from the professional staff to computer processing. Computer editing of direct structural input is well developed, and the development of computer editing techniques for nomenclature has begun. Computer-based editing of nonbibliographic, nonstructural information lies in the future.

5. Development of computer file organization and accessing techniques that will assure a responsive, reliable, and economical system with long-range viability. A great amount of skill is needed in coping with the problem of building large computer-based information files and of maintaining and using them individually and in combination. The interlinkage of CAS computer files will have to adjust as the files grow if the system is to remain responsive and economically viable. The manner in which the system is used by the community will also deeply affect file organization developments.

6. Close interlinkage of CAS activities with those of other primary publishers, secondary information services, and resource libraries.

7. Development of effective means of supplying information to information centers for redistribution. The interfaces in this area include the development of data content and format standards for information exchange and the determination of appropriate media for that exchange. Since much information exchange is based on the documents in the system, a document registration system is being considered. CAS is also working

on algorithms for the unambiguous identification of documents (including both serials and individual articles within serials).

Computer-Based Services

CAS now offers the following computer-based services on magnetic tape in a form that permits searching of the titles, abstracts, or index terms contained in each issue:

- *Chemical Titles*--a bi-weekly alert to the titles of new papers appearing in 650 journals of pure and applied chemistry and chemical engineering.

- *Chemical-Biological Activities*--a computer-produced abstracting and indexing service in biochemistry covering 15,000 abstracts per year, dealing with organic compounds and their effects on biological systems.

- *Polymer Science & Technology*--a weekly, computer-based guide to the current journals, reports, and patent literature of the chemistry, chemical engineering, and technology of polymers.

- *CA Condensates*--weekly, computer-searchable tapes containing complete abstract headings and *CA* issue keyword index terms for every journal article, report, and patent covered in the corresponding weekly issue of *CA*.

- *Basic Journal Abstracts*--bi-weekly, machine-language tapes containing the complete *CA* abstracts for all articles appearing in 35 chemical journals. Headings and full texts of abstracts can be searched by computer to derive a bibliography of articles of interest to the searcher. Printed copies of abstracts accompany the tapes.

Relations With Other Centers

Output centers using CAS computer-generated materials are operating in five U. S. university centers and in Canada, Denmark, Sweden, and England. The Netherlands and Belgium are expected to be participating soon.

International cooperation in chemical information systems is being developed through agreement between the ACS and the United Kingdom Consortium for Chemical Information, of which The Chemical Society (London) is acting as the contracting party, and between ACS and the Gesellschaft Deutscher Chemiker of Germany. Each of the European organizations will assume responsibility for marketing in their countries the publications and the computer-based information services of CAS.

Both are working with CAS in preparation for abstracting the chemical literature of their countries for input to the CAS system. Discussions with the Japanese have begun, to consider the possibilities of a third such international agreement.

Other Dissemination Media

All ACS journals now are on microfilm that is leased to users who pay license fees per original page; the fee permits the lessee to make hard copies to the extent that he wishes for use within his own organization. Microfiche also is being examined for various uses, including the recording of experimental details and data that may not be included in the pages of published journals.

Audio and video tapes offer promising possibilities for information handling. ACS is developing for delivery in January 1970 an audionews system that will furnish 15-minute cassettes weekly to subscribers. While the first of these units will be devoted primarily to technical and economic material, an audionews unit for research and development is under consideration.

H. WILLIAM KOCH

American Institute of Physics
New York, N. Y.

USER ORIENTATION OF PRIMARY JOURNALS IN PHYSICS

The principal communication mechanism in the discipline of physics, the primary research journal, is being redesigned. Redesign is made necessary by the growing problems of size, cost, and time delays that are outmoding all conventional mechanisms such as primary, abstract, and titles journals. Redesign is made possible by sympathetic, but pressing, national support and by new technologies. This report focuses on preliminary plans to supplement the primary journal system in physics with a user-oriented set of journals (or their equivalents) that will permit primary journals to continue to serve a unique but modernized role. These plans are not final at this time and are dependent on financial support from the National Science Foundation as well as on approval by the AIP Governing Board.

Growing Problems with Primary Journals

AIP's position as publisher of about 35 percent of the world's literature in physics leads us to anticipate that the total stack height of primary research journals in physics for the world in 1969 will be about 60 feet. Let us examine the implications of the 60-foot stack by comparing it with the information problems of physicists in 1940. A volume of *The Physical Review* for the entire year of 1940 equalled in thickness that for one month of today's production. In 1940 physicists had little difficulty in monitoring the research completed or in progress anywhere in the world in their subdiscipline. In 1940 most physicists could and did subscribe to *The Physical Review* or other physics journals in which their kind of physics was likely to be published. In 1940, individual recording and use of material from physics journals consisted of typing, photographing, or handwriting excerpts from the journal pages.

Not so today. Today the proportion of physicists that subscribe to *The Physical Review* is decreasing. The reasons are obvious. The first is one of sheer bulk. Even though some physicists' offices are adequate to house 5-foot-wide annual productions of the *Review,* the realization that a 15 percent growth rate implies a doubling in 7 to 8 years is discouraging to individual physicists. The second reason is cost. Even though the *Review* and other physics journals published by AIP are inexpensively produced and are available at low subscription prices per page (average nonmember price for AIP journals is about 0.7 cents per page), the total number of pages is going up. Thus, when some single issues of such journals as the *Review* or the *Journal of Applied Physics* reach 700 pages, the issue must be sold to nonmembers for almost $5, a substantial sum.

Another reason is relevance of the material in any one journal issue. If the material that is directly relevant to an individual physicist consists of 10 or 20 pages per issue, he is better off financially making photocopies even at 10¢ per page, thereby producing his own customized *separates* journal. If, in addition, because of the spreading geographic as well as the growing interdisciplinary nature of science, he realizes that the physics in which he is interested is no longer limited to a single journal, but may appear in any one of 70 journals published by foreign as well as U. S. publishers, commercial as well as society, then he discards the convenience of his own personal office copy and relies increasingly on library copies plus photocopying.

Let us speculate on the future of the conventional journal in physics if no changes are instituted. Journals are likely to become library editions and reference works in the true tradition of an archive. This bodes for an uncoupled, massive, antiquated communication system that will collapse of its own weight, unless steps are taken to redesign the journal system to be a more efficient communications medium. Otherwise, in desperation, much less efficient and less responsive schemes such as preprints will be invented to satisfy the needs of small isolated components of the physics user community. The future is forcing on us the development of a coupled, modernized information system which, in my opinion, must be built on the important foundation of a revitalized journal system.

Analysis of Problems

The basis of science information problems is obviously the well-publicized exponential growth of scientific and technical literature. The present doubling time of 7 1/2 years for physics literature means a particularly severe problem for the discipline of physics. In 1968 alone, over 50,000 research articles in physics were published throughout the world in about 500 journals; over one-third of these appeared in the 34 archival and translation journals published by the AIP.

The seriousness of the problems for physicists can be appreciated from the implications of the numbers in Table 1. The numbers show that the members of the AIP publish in AIP journals about 4 times as many articles per year as the members of the American Chemical Society publish in their journals, and similarly for members of the IEEE. These numbers are not conclusive, but they do suggest that physicists publish more in, and depend more on, primary journals than do chemists or electrical engineers. Therefore, when the journals start having problems, physicists become particularly concerned and sensitive.

TABLE 1. 1968 MEMBERSHIP AND JOURNAL DATA

	AIP	ACS	IEEE
Membership	48,683	126,269	186,298
Primary Journals	21	20	36
Editorial Pages	65,887	40,225	23,759
Approx. Pages per Member	4/3	1/3	1/8

The exponential growth is blamed by some on the growing pollution of the literature. This would not appear to be the case in physics, since the number of pages published by member physicists in AIP journals has been a reasonable constant for about a decade. Therefore, the growth in the literature can be attributed to the growing number of physicists who publish annually and not to their increasing productivity (and any concomitant decrease in quality).

The expanding volume is flooding even the narrowest specialties, making it difficult for the individual physicist to keep up with areas of immediate interest to him, let alone those bordering areas and new specialities with which he should be familiar. The situation is aggravated by the producer orientation of the present journal system in that the author largely decides when, where, and in what manner the information is to be presented. The user is left to cope with the flood as best he can, often resulting in information coming to his attention too late to be of use.

User Orientation as One Solution

One of the quickest and most expedient ways of solving the growing information problems of the physics community is to include in the design and operation of the primary journal system the needs of the users as well as of the producers. The orientation can be accomplished by producing a completely new set of user physics journals resulting from a manipulation of the articles in the archival physics journals. The content of the user journals would be determined by standard interest profiles of the scientists and engineers who would use the articles in the journals.

The possibility and practicality of producing user journals in physics through the AIP results from the relatively small number of U. S. physicists (32,000), from their inclusiveness in AIP membership (about 75 percent), and from the concentration in publishing at AIP

(35 percent). This possibility can be pursued on a trial basis with AIP literature first. After demonstrating to ourselves and to commercial and foreign publishers that user journals will strengthen the journals of the individual participants and will make possible the continued contribution of journals as the fundamental science communication mechanism, we can then expand beyond AIP literature to include the world's core literature in physics.

Let us now become specific about user journals in physics and the future information system trends at AIP. In the past, we have owned and operated a set of journals for our membership; these journals satisfied their needs as users as well as authors (producers). In the future, the operation of the archival journals for our members as authors will continue to be necessary. However, as users they are increasingly interested in obtaining rapid and economical research results regardless of what geographical location of the world or what scientific or engineering discipline was the production source of the information. Therefore, AIP plans to expand its marketing of journals beyond the 21 journals published by us. We plan to market at least 13 translation journals, the eight journals of the Institute of Physics and the Physical Society (London), and a few other foreign journals. The marketing will include eventually bilateral agreements and arrangements on the handling of hard-copy, microform, bibliographic material on computer tape, indexing, and integration into what has been called the National Information System for Physics (NISP). It is our hope that we can go beyond the total of 53 journals now marketed to include in NISP substantially most of the world's physics literature.

One result of NISP will be the production of the user journals or their equivalent. For the purposes of illustration, let us assume that AIP will make available 300 user journals with journal contents determined by standard interest profiles modified annually. The user journals will select material from the archival journals slightly before the time of publication of the latter. Whereas the articles in the archival journals appear only once, a given article would appear in as many users' journals as would be required. Articles of broad interest would most likely appear in several user journals.

The particular formats of the user-oriented services and combinations thereof that will find user acceptance will have to be determined by market surveys and pilot tests. Those now being studied include: 1) Selective notification of information, a listing of articles in the journals during a given time period that match the standard profiles of the 300 or so special interest groups in physics; 2) Packages of microfiche, selection of articles in microfiche form being sent to subscribers at periodic intervals; 3) Cartridge of microfilm, selection of articles copied onto a 16-mm. strip and mailed in a cartridge to subscribers; and 4) Hard copy selection, user journals published from offprints of the original journals.

Emphasis in the production of the above sets of user journals or their equivalents would be on quality, timeliness, and economy with the objective of giving to physicists and other users a customized service superior to their own (haphazard) collecting and photocopying. It is our goal to make articles available simultaneously in the user and in the archival journals. It is also our goal to integrate the concept of the user journal into the over-all National Information System for Physics with its long-range goals of a sophisticated (interactive) communication system that will benefit all scientific and technical users of physics information.

MELVIN KRANZBERG

Case Western Reserve University
Division of Special Interdisciplinary Studies
Cleveland, Ohio

THE TECHNOLOGY TRANSFER ANALOG

As an historian of technology, I make one basic assumption in approaching the problem of improving the communication of scientific and technical information to its users. This assumption is that technology is basically an information system, no matter whether we are dealing with machines, tools, processes, the organization of work, or any other of the myriad elements involved in technology. If technology can be viewed as an information system, then the process of transferring technology can be viewed as a special case of the dissemination of information. In brief, the transfer of technology might serve as an analog for the problem of communicating scientific information.

The transfer of technology, like most problems connected with the transfer of information, has many different aspects. First, of course, is the transfer of information from one generation to the next. Since this is the entire question of education viewed as a device for transmitting man's social inheritance and accumulated knowledge to posterity, it is a problem also affecting science, literature, laws, attitudes, values--indeed, all elements of our culture.

Second, in the context of the national economy, the problem of technology transfer applies to the transference of tools, techniques, machines, processes, etc. from an initiating firm to other firms in the same industry. Another category is the transfer of such items from one technology to another technology, as, for example, transfer of the techniques of air-frame manufacture to the construction of prefabricated housing.

One can also look at the dissemination of technological information in an international context. One category is technology transfer from one industrialized country to another. Another involves the more difficult but more pressing problem of transfer from industrially developed to underdeveloped countries. This is a transfer between cultures, one at a sophisticated level of technology to one at a more primitive level.

Finally, there is an entirely different aspect to this problem which does not involve technology transfer per se. It is the information dissemination problem from basic science to technological application. This major problem has been catapulted into even greater significance, perhaps unwarrantedly so, by the common view that technology is simply applied science. If that is so, the chief problem which faces us is to acquaint technologists with the results of the latest scientific research so that they can immediately proceed to apply them. Since, however, technology is much more than, and sometimes quite different from, the direct application of pure scientific research, this problem is more complicated than it might appear.

Where do we stand in our knowledge of the dissemination of information, whether considered as education, as the transfer of technology, or as the application of science? The answer, unfortunately, is that our knowledge lags far behind our intentions. In the application of basic scientific research to technology, for example, historical scholars cannot agree even on the meaning of those few specific cases that have undergone thorough study, let alone arrive at generalizations having wide application. There has, for example, been much study of the role of science and technology in James Watt's invention of the steam engine, some scholars claiming that Watt was led to his invention by Joseph Black's researches on specific and latent heat, while others claim that Watt invented the separate condenser empirically and without any basic understanding of heat phenomena.

Helping to mislead us on the question of science-technology relationships are charts showing increasingly swift application of basic scientific discoveries. These charts indicate that it took 1600 years from Hero's aelopile to Watt's steam engine, 25 years from Faraday to the electric generator, and about five years from Bardeen and Shockley's research to the commercial use of transistors. Extrapolation suggests that by the year 2025, it will take milliseconds to go from enunciation of a basic physical theory to its practical application.

The trouble with such charts, of course, is that they rely on selective evidence. Judicious selection of other bits of evidence could perhaps prove that the time between basic research and application is lengthening rather than diminishing. Additionally, such charts sometimes depend on very tenuous historical connections. Watt's steam engine did not come from Hero's aelopile but from Newcomen's atmospheric engine of 60 years earlier, and the Newcomen engine was an outgrowth of interest in pressures deriving from the preceding 17th century. On the other hand, work on the theory of the transistor can be traced back further than Bardeen and Shockley, to the 1920's and earlier in this century.

What we do know historically about science-technology relations would seem to indicate that, for the most part, new technology comes chiefly from old technology without much relationship to the latest scientific discoveries. If technology today seems more reliant upon science than in previous times, perhaps it is because technology has itself become more of a science than an empirical, craft art. The clues to the relationships between science and technology are thus not to be found in any simple transference of scientific data to the technological community but must be sought in more circuitous routes. This might lead us to redefine the problem. Perhaps it is not a question of getting the latest *pure* science into the hands of technologists but getting the latest science-based technology diffused to technologists. Solutions to this problem might lie in changes in engineering education, in the organization of industrial research teams, in the patent system, in professional societies, and in engineering publications.

In developing an analog of technology transfer, one finds few historical studies leading to meaningful generalizations. These studies have only recently gained prominence, partly because of attempts of NASA and the Department of Defense to justify their research expenditures and of attempts to advance the technological level of industrially backward nations. One central fact of historical studies has emerged, namely, the prime importance of the person-to-person quotient, or, more simply, the role of people in the process of information dissemination.

History shows, for example, that American technology during the colonial period relied almost entirely on British technology brought by immigrant craftsmen from their homeland. This continued during industrialization in the first half of the 19th Century when the U. S. simply imported British machines and tools and the workmen who knew how to operate them. Alternatively, Americans went abroad to learn about the machines and returned with this knowledge and the machines. Today, we are exporters of technology but we find it necessary to export, together with machines and processes, the operating know-how that must be conveyed by people.

An eminently successful illustration of the effect of person-to-person transfer has occurred in agriculture. Productivity of American agriculture has soared since the beginning of the 20th Century at a rate outstripping that of improvement in industrial production, although more attention has been given to industrial phenomena such as the assembly line and automation. How was the agricultural revolution accomplished? Textbooks attribute the sharp rise in productivity to the development of scientific agronomy, the use of machinery, and the application of capitalistic methods to farming. True, but how were scientific methods made known to the individual farmer stuck out in the country? How did he learn about new strains of hybrid corn and wheat, the fertilizers, and pesticides which could multiply his yield? Where did he find out about the machinery and how to operate it? And now did the farmer learn to apply industrial techniques?

First we must ask where the knowledge originated. The original research had two sources: 1) educational institutions, mainly state agricultural colleges which are land-grant institutions; and 2) laboratories, mainly those of the U. S. Department of Agriculture and of the state agricultural experiment stations. The colleges transmitted some of the information to the newer generation of farmers through educational processes. However, the bulk of continuous dissemination was carried by state agricultural extension services and county farm agents who brought new knowledge directly to the farmer on the farm. There were, of course, other means of research dissemination--farm journals, county fairs, and demonstration projects, some of the latter effected by commercial seed producers and manufacturers of farm equipment. These transfer agents also relied largely on personal transfer of knowledge.

The effectiveness of such means of disseminating research results to the user is also demonstrated by their success abroad. Dramatic increases in rice and wheat production in food-hungry countries resulted from the application of farm extension services to these nations by which people were taught how to use fertilizers, pesticides, and farm machinery and how to plant, cultivate, harvest, and store.

Is, however, the example of transfer of agricultural technology pertinent to questions of the transfer of scientific research data within the scientific community? To a certain extent, scientists are used to obtaining information through reading, and demonstration and personal communication are not essential. One might note, however, that the community has sought interpersonal communication since the time of Plato's Academy and the Museum of Alexandria. At the time of the Scientific Revolution, personal communication became institutionalized with the formation of the Royal Society and the Academie des Sciences. Today the flood of publications makes it almost impossible for the scientist to read everything in his own field, and he has resorted to invisible colleges and conferences to help him, in a sense, separate the wheat from the chaff and to keep himself up to date. Additionally, growth of postdoctoral research projects, exchange professorships, and international scholarships for training scientists of backward countries in the schools and laboratories of developed nations testify to the need for personal interaction. A tendency to idealize and depersonalize science and technology overlooks the fact that these are human endeavors in which personal exchange can be more efficient than impersonal mechanisms.

Nevertheless, even institutionalized modes of personal contact cannot cope with all of the information dissemination requirements. Some way must be found to store data, to retrieve the useful data, and to bring that scientific knowledge to the individual who can apply it fruitfully. If we accept the idea that the transfer of knowledge requires people as well as publication, we appear to need a new breed of middleman, an information-transfer specialist, whose particular task would be to bring the results of research to its possible users. One model of this specialist is the entrepreneur in technological enterprises. He brings together the basic technical idea, the business expertise, the economic demand, and the capital necessary to translate an ingenious idea to a profitable innovation. Many an invention would have been stillborn but for the entrepreneur who saw its possibilities and successfully brought together the many elements required to convert it to practical application. Watt's steam engine, for example, probably needed the entrepreneurial work of Matthew Boulton. Similarly, Edison would probably not be remembered as America's outstanding inventor had he not combined entrepreneurial qualities with his inventive genius.

The research manager in large-scale industrial research laboratories is sometimes like the commercial entrepreneur. He brings together a team possessing scientific knowledge and technical expertise, adding his own knowledge of the economic constraints and the social and economic needs. In a sense the successful research manager works as an agent to disseminate knowledge from many different fields to produce profitable applications.

It seems worth investigating the characteristics that could be institutionally developed in people for the role of information-transfer specialist or *information entrepreneur*. However, we must avoid the mistake often made by professional educators who become so infatuated with techniques (for example, the techniques of information dissemination) that they lose sight of the substance to be taught. Good farm agents very likely know a great deal more about farming than they do about the techniques of information transfer. Perhaps we should take our clue for solving the problem of communicating scientific and technical information from those who have already done it successfully.

JEFFREY NORTON

CCM Information Sciences, Inc.
New York, N. Y.

COMMERCIAL INFORMATION PRODUCTS AND SERVICES

CCM Information Corp. is a wholly owned subsidiary of Crowell Collier and Macmillan, now the largest book and general education publisher in the country. Our subsidiary is a publisher of books, journals, and information services in a variety of subject fields. Created primarily to fill information needs that can best be served through nontraditional publishing methods, our services stress effective use of computer processing and microfilm techniques.

Much of our product is published in several media, depending on customer requirements. Not infrequently we publish a given data base in print, microfiche, microfilm, and computer-tape form, and will also offer special search services based on access to our computer. Our bibliographic computer files are all based on the Library of Congress MARC II format to facilitate use by libraries with their own computer-based systems.

Our long-range objectives include not only the publishing of catalogs and definitive indexes to important information files but also, whenever possible, the publication or distribution of the substance of what has been indexed.

Products and Services

The *Pandex Current Index to Scientific and Technical Literature* is a data base that stores, annually, title and author information on approximately 250,000 journal articles, 35,000 *USGRDR* technical reports, and 6000 new books. It is primarily a current-awareness service through which information is provided on published literature, frequently within 12 days of publication of the original source document. All works are indexed under an average of six computer-generated and thesaurus-controlled subject headings, as well as by author. The data base is available to customers weekly in magnetic tape form, biweekly in printed format (about 400 pages per issue), and quarterly and annually in microfiche.

Three separate publications have been developed from our new computer-based information file, *Current Index to Conference Papers*. Here the emphasis is on speed of processing to provide up-to-date information quickly, frequently before some of the cited papers have even been delivered.

The quarterly publication, *World Meetings*, provides information two years in advance on important scientific and technical meetings to be held throughout the world.

Calls for Papers is a weekly alerting service that lists all "Calls" for scientific and technical meetings in a standardized and reproducible format.

Research and Microfilm Publications (RMP) provides monthly indexes, with corresponding microfilm or microfiche, to all the translations (about 250,000 pages per year) of the U. S. Joint Publications Research Service. Even though all documents are also available from the Clearinghouse for Federal Scientific and Technical Information, our indexed and packaged service sells widely to research libraries and government agencies in both the U. S. and abroad. Until recently, the *RMP* service was organized and indexed solely on a geographic basis. This fall, we started a new service that indexes separately all scientific and technical material, regardless of country of origin.

Current Index to Journals in Education, inaugurated in 1969, is the result of an innovative partnership between the ERIC network of the U. S. Office of Education and ourselves as representative of the private sector. The program indexes the newly published educational serial literature and includes a monthly journal, semi-annual and annual hardbound cumulations, full computer-tape publication, and plans for a broad spectrum of specialized subsidiary publications. The publication contract was let under standard competitive-bidding procedures and embodies the concept of a phased transfer over a period of years, from partial government subsidy to complete support by the publisher.

Through partnership with a London-based firm, Information Retrieval Ltd., we offer *Biomedical Abstracts,* in seven subject areas: Microbiology, Virology, Genetics, Aquatic Biology, Calcified Tissues, and Entomology. Publication is monthly with an annual index.

Our National Auxiliary Publication Service is a unique microfiche publication program operated in conjunction with the American Society for Information Science. Its purpose is to make available in microfiche important material supplementary to basic articles published in the professional journal literature. Supplementary material is received from journal editors or authors, microfilmed, indexed, and assigned an accession number. This information is printed in the journal with the summary article, and interested parties can obtain the supplementary data for the modest fee of $1 per microfiche.

New Approaches

We feel that the commercial information industry has embarked on several programs that either offer or promise progress in scientific publication. In some cases, this progress is conceptual; in others, it is more simply the belated application of private sector initiative to problems that are as old as awareness of the information explosion itself.

A notable example is the *Current Index to Conference Papers* data base. The fact that it is published not only in print but in weekly computer-tape form in a standard format is something of an innovation, and hopefully represents publication progress. Much more significant to the future of scientific and technical publications, though, is the basic fact that for the first time, information on the conference-paper literature from over 2000 worldwide sci/tech meetings per year is being made widely available. Each year we expect to index about 150,000 papers. Based on current statistics, about 80 percent of these will eventually be published in the professional literature, but publication frequently won't occur for a year or two or even longer.

Recent research at the Center for Scientific Communication at Johns Hopkins has established that, for active research scientists and engineers, the papers delivered at conferences are a more important source of information than the published literature. *Current Index to Conference Papers* may introduce a demand that could provide the impetus for dramatic changes in scientific and technical publication practices.

Instead of the 8 or 10 papers now provided monthly by the typical professional journal, why not have one that would offer abstracts of 40 or 50, and include the full text of each on microfiche in a pocket on the inside of the back cover of the publication? Outwardly, the publication would appear much like those produced now; in print would be news articles, letters to the editor, the 40 or 50 abstracts, and of course sufficient pages of advertising, yet information on important research in the field would be reaching subscribers at a four- or five-fold accelerated rate.

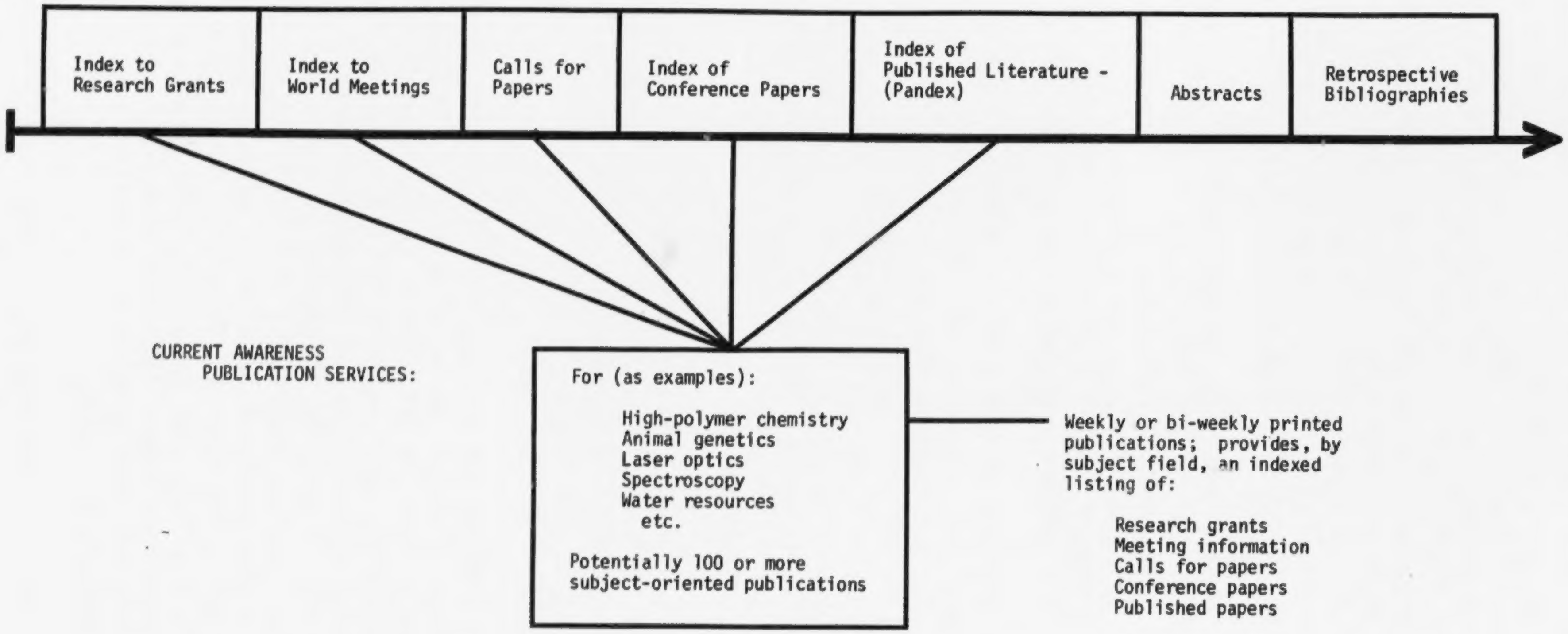
Early next year, we hope to initiate a program that will offer conference papers on microfiche through arrangements with sponsoring societies or authors. Audio tapes of key papers may also be offered. Having created the index, we now view our next job to be providing the substance itself.

Another innovation in our program to provide secondary information about research in science and technology involves our plans to offer specialized subject-field current-awareness services. The diagram on the following page represents a time-scale continuum of public information about scientific and technical research. For information to be most useful, it must be organized to the needs of the individual. Thus our planned Current Awareness Publication Services (CAPS) are designed to provide the research specialist with all the secondary information he needs about who may be, is doing, or has published what. We feel these CAPS services may be a more practical way of filling sci/tech current awareness needs than individual SDI services that are expensive now and offer little promise of a marked drop in price even with increased numbers of users.

Other plans include enlarging our coverage of the foreign literature and making primary data files available and usable in computer-tape form. We believe tremendous opportunities for progress exist through constructive associations between government and the private

TIME-SCALE CHART OF INFORMATION ABOUT RESEARCH IN SCIENCE AND TECHNOLOGY

(Plan for a Comprehensive Information Service)


Index to Research Grants
Index to World Meetings
Calls for Papers
Index of Conference Papers
Index of Published Literature - (Pandex)
Abstracts
Retrospective Bibliographies
CURRENT AWARENESS PUBLICATION SERVICES:
For (as examples):
High-polymer chemistry
Animal genetics
Laser optics
Spectroscopy
Water resources
etc.
Potentially 100 or more subject-oriented publications
Weekly or bi-weekly printed publications; provides, by subject field, an indexed listing of:
Research grants
Meeting information
Calls for papers
Conference papers
Published papers

sector. New information systems frequently need massive investment to get them underway, not uncommonly an investment greater than can be undertaken by a single commercial organization. But the information system, after it has been created, can profit from the publisher's art--packaging, advertising, selling, and customer education. I can only hope that programs similar to the one we have underway with the Office of Education will be developed by other government agencies.

SIDNEY PASSMAN

United States Arms Control and Disarmament Agency
Science & Technology Bureau
Washington, D. C.

SOME CONSIDERATIONS CONCERNING IMPEDANCES IN THE SCIENTIFIC COMMUNICATION NETWORK

We assume that the forcing function of the national research and development effort stems from the demands of society for satisfaction of its felt needs towards achieving better health, security and material well-being, and that these desires will continue unabated in the foreseeable future.

In the final analysis, the nation's effort is kept from the greatly increased productivity that society hopes for, for a number of reasons. Basically, the R and D process is a creative one, admittedly not too well understood. It is hard to accelerate its accomplishments by simple additions of manpower and other resources. The secrets of nature and the elements of discovery and inventiveness apparently cannot be forced.

In addition, the scientific communication system, which serves to provide the important multiplier effect in R and D by permitting workers to profit from the knowledge and experience of co-workers, contains impedances in its networks, some of which may have hard cores of inertia which will be extremely difficult to remove,[1] and others whose amelioration is possible but which will require much greater concern and effort on the part of all involved.

I will simply outline the latter elements and then discuss them in detail sequentially. In particular, the role of the Federal Government, as a principal agent of responsibility for society in this endeavor, will be discussed.

Topics of Relevance to the Impedance Issue

a. The *business aspects* of R and D.

   1. proprietary issues

---

[1] The engineer will recognize that the removal of all impedances in these positive feedback loops would lead to a divergence or true explosion in the overall effort. One can be consoled against this happening by recognizing that as long as people are involved in the effort, there will always be some inertias.

    2. competitive nature of industrial activity
    3. commercial issues in information handling
    4. *property* aspects of information
    5. information as a resource in the R and D process

b. Secrecy, national security, and international rivalry.
    1. *need-to-know*!?
    2. proper balance
    3. one-world of science vs nationalism

c. Language problems.

d. Mores of the scientific community.

e. Time delays in various aspects of the communication problem.

f. Human factors, e.g., informal practices, limitations of the individual - in assimilation, etc.

Business Aspects of R and D

No one can really understand the scientific communication process unless he first appreciates that in recent times we and other nations have adopted a very pragmatic view towards our scientific and technical establishments. Thus science and its twin, technology, have become a big and vital *business*.

1. It is clearly recognized that scientific information plays a significant role in this business process and for that reason it is treated as *proprietary*. The fact, however, that so much of the R and D process is sponsored or supported by public funds demands that its results be placed in the public domain as soon as possible.

This area is close to that of Patent Policy and must be handled in concert with it (cf. the recent FCST recommendations on this subject); however, it should be pointed out that the fraction of important new scientific information that is being treated as proprietary is increasing. Its expeditious release must be insisted upon by government.

2. In spite of the fact that U. S. industry is considered to be "all on the same team" with regard to major national undertakings like the Defense program, the Space program, H.E.W., etc.--we operate through a competitive industrial base. An internal data base of scientific information and *know how* is often considered a principal asset of one group in competing with its competitors. To that end, crucial information is often withheld or *written around* in scientific papers or technical reports. This situation can only be changed by conscientious project officers who seek to improve the quality and completeness of the reports issued on their programs.

3. While it has become recognized that the communication efforts should be considered as elements in a *system*, it is well known that various portions of the effort are *managed* by different groups with different motivations and practices. A most important portion, including valuable primary and secondary materials has come under the control of commercial interests. While recognizing the free enterprise assets as important it must still be the concern of the scientific community and the government to oversee the performance of this portion of the system. Both the elements of journals and books are too important in the overall system to be allowed to deteriorate in any way without interfering with the integrity of the system.

Likewise, concern for the viability of other aspects of the private sector performance must be observed.

4. Information is intellectual property. Its proper protection (copyright, etc.) serves as an incentive to further efforts in the field. At the same time advances in reprography and machine handling of information permit new improved ways of exploitation and personal access to information. A fair solution must be found in the immediate future to the proper economic and incentive support and use of intellectual property.

5. Information is a resource to the R and D process. The proper financial support to this resource and its relative priority in consideration with other resources must be further addressed. Furthermore, the measures of effectiveness for alternative elements in the communication process deserves increased attention.

Secrecy

Compartmentalization and need-to-know restrictions are instituted to purposely restrict information flow to avoid compromise of information to the *enemy*. Naturally they serve to keep information from other programs in our own efforts which might profit from this knowledge, including other governmental programs and other national goals.

There exist no opportunities to judge the relative merits of wider dissemination other than within individual Agencies of government. Some eminent persons have called for a "wider forum" in which to consider these issues. I consider this an extremely important issue which so far has apparently escaped the concern of the Chief Executive's staff.

In addition the *integrity of science* and effective *technology assessment* are severely handicapped by these restrictions. The recent report of M. Mead's AAAS Panel and its predecessor group are worth noting in this regard.

JACK SIEBURG

Office of the Judge Advocate General, USAF
LITE Activity
Denver, Colo.

INFORMATION POTENTIALS OF FULL-TEXT COMPUTERIZED RECORDS

Many approaches are being studied today to determine how best to take advantage of new information storage and manipulation devices. At LITE (Legal Information Thru Electronics), we convert the full text of documents to machine-sensible media, that is, we store every word of the original body of text material. The current LITE master file consists of approximately 75 million words of text and includes such major texts as the *United States Code*, the published and unpublished *Decisions of the Comptroller General*, and the *Armed Services Procurement Regulations*. Our design philosophy is that, with full text, we have virtually unlimited possibilities for generating information product subsets and management aids.

The LITE system presently operates under contract on an RCA Spectra 70/45. Peripheral equipment includes a high-speed printer, card reader and punch, 800-bpi magnetic tape drives, and disk and magnetic card immediate access storage media. All text and word locator files are stored on magnetic tape for processing, under a tape-oriented search and retrieval system, and for emergency backup purposes. Files considered to have major research value or that are subjected to a high rate of activity are also stored on immediate access devices and processed under a random-access system.

For search and retrieval operations, LITE uses an inverted file of *word locators* which identify the exact location of every nontrivial word in a particular body of text. Word locators are not developed for common or subjective words such as *the, and, but,* but their position in the text is accounted for in developing the locators for the noncommon words. The word locator information is contained in a computer-generated 12-byte fixed-field code.

LITE programs incorporate several techniques for restricting the amount of output. Among these are limitations that 1) require that specified word relationships occur only in (or do not occur in) certain types of text, 2) documents identified for retrieval are current and in effect, or 3) were effective during specified time spans. The first capability results from the use of *type codes* and include/exclude type instructions. Control data for the last two techniques are contained in a machine-generated control record that precedes each text document.

Documents identified for retrieval are available in one, or a combination, of three types of printed output: context, citation, or full text. The first, context, produces three lines of text showing how the search words were used in context, the words requested appearing in the middle line. The citation output tells where, in what specified document, the words requested may be found; for example, the page within a volume of *Decisions*. The last output produces a print of the entire text; normally, due to print-time factors, this method is not used if the person requesting the search has access to the printed material.

LITE file maintenance operations cover two distinct types of data banks, those that are updated by adding new material to the end of the present file, and those that must be updated by the insertion of revised or new material. For the former type files, new material is assigned document numbers in ascending sequence and the master file grows until it reaches a workable limit of approximately 7.5 million words; at this point, a new file is started. For files that require revisions, i.e., the *U. S. Code,* three techniques may be employed. The file maintenance techniques for updating the master text files also simplify updating of the master word locator files. Word locators are developed for the new material and are then merged with those developed for the old master file.

Specialized indexes prepared by the LITE system have proved extremely valuable. The term *selective* in Selective Key Word In Context Indexes may refer, in the LITE system, to keywords (words for which index material will be developed) or stop words (words to be excluded from the final index). The final index print routine can be directed to delete additional stop words and/or phrases. The LITE system also has the capability of providing index material on magnetic tape, properly formatted for photocomposition purposes. Formatting capabilities include the use of light and bold-face type, indentation, use of the leading factor and, in some instances, the inclusion of upper/lower case.

Another type of specialized index is the cross-reference or citation index. In the development of this type of index, one body of text is electronically processed to locate all references (citations) to another specific body of text (e.g., citations to the Court of Claims cases appearing in the *Decisions of the Comptroller General*), or to another section of the same body of text (e.g., references in Title 10 of the *U. S. Code* to Title 5 of the *Code*). The printed index may be sequenced in either citing or cited reference order.

Another special index developed by LITE identifies all *Decisions of the Comptroller General* that have been overruled or otherwise modified. The system, consisting of three separate phases or programs, takes advantage of the man-machine interface. Computer analysis first locates all decisions that have been referenced by subsequent decisions, the original computer output consisting of data identifying the modified and modifying decisions, in both printed and card format, plus a contextual print of the modification action. Then, the type of modification action (overruled, modified, distinguished, etc.) is entered by people, appropriate codes are gang-punched into the data cards, and the reference card deck is processed to magnetic tape and sorted in ascending sequence on the modified decision reference factor. The final index is then formatted and printed out.

One of the special products available from LITE is text editing. The original text editing routine was designed specifically to assist management in the complete revision of an existing regulation. Prior to preparing a draft of the revision, the LITE staff designed and developed a computer application that 1) locates and identifies (underscores) all suspect words and 2) lists the proposed substitute word or words in the right-hand margin. (A suspect is a word that may be superfluous, ambiguous, or for which an authorized abbreviation might be substituted). Use of this editing technique in the revision of the regulation resulted in a size reduction of approximately 40 percent, improved clarity, and a higher degree of standardization.

Future plans for the LITE system include a redesign 1) to improve and expand search and retrieval operations and 2) to incorporate remote access capabilities and reformatting of files. The use of microform for storage of data bases and index data is being studied. Present plans envisage copies of the files being available at regional sites with the required scanning and output media; index data could thus be available at a much reduced cost with a minimum of storage requirements. Transmission of information over telephone lines is also being studied.

PHILIP J. SIEGMANN

Georgia Institute of Technology
Atlanta, Ga.
(formerly with Amer. Psychological Assn.)

EXPLOITATION OF FILES FOR INFORMATION PRODUCTS

The purpose of this paper is to consider some of the products and services that can be derived from machine-readable and other files such as are generated by discipline-oriented abstracting and indexing services (hereafter abbreviated A/I services). The products and services that can be produced from presently available files will be considered first. Next, the use of personnel directories and other files that can be merged with A/I files to produce new services will be discussed. Finally, some suggestions will be made for the inclusion of additional data elements in A/I files that promise to increase their utility as a source of new services. The time scale is the near future, and most of the products are printed publications or computer printouts.

A study by Herner & Co. of the *Psychological Abstracts* file considered economic factors and probable markets in the development of one representative list of such products. The following were identified as available from machine-readable abstracts and indexes: 1) cumulative indexes, 2) computer tape copies of the A/I file, 3) special interest subsets of the A/I publication, 4) continuing bibliographies for mission or multidiscipline oriented users, 5) topical bibliographies, 6) indexes for primary publications, and 7) directories of research activities. Of the foregoing, only the last two can be considered innovative or unusual. Since these products are *conventional*, however, one can extrapolate to their use with some confidence.

Based on an hypothesis that local directories of research have value to researchers in a given geographic area, a list of entries in *Psychological Abstracts* that emanated from Washington, D. C., institutions was compiled. Questionnaires were sent to local authors and responses were used to supplement the computer-generated information. A directory was printed and distributed for evaluation purposes. Responses indicated that local directories might well serve some information needs. Our experience suggests that useful directories can be generated directly from A/I files. It is important to note, however, that a small percentage of the professionals in a field produce most of the scientific literature. Thus, while there are approximately 30,000 psychologists in the U. S., less than 10 percent produce between 70 and 80 percent of the literature. Similarly, a limited number of institutions are sources of most of the published research, and a limited (approximately 20 percent) number of the journals provide most of the relevant materials. Furthermore, there is a substantial correlation between the most productive authors and the most productive institutions that, in turn, tend to publish in the core journals of the discipline.

The core file for psychology will thus contain information on about 2000 to 2500 authors, 50 institutions, and 200 journals. This can no doubt be generalized to other fields.

This suggests that a comparative analysis of an author file with the classification schedule in a field would provide basic information on the major areas of a discipline in terms of its current literature. For example, there appear to be approximately 120 principal areas in psychology. An innovative use of the directory file was in the planning and programming of scientific meetings and conventions.

Surveys of the users of A/I services have frequently indicated that users have difficulty in obtaining noncore, or peripheral, literature, particularly foreign technical reports and monographs. Since A/I services maintain current addresses for sources of technical reports and monographs, and less well known journals, this information could be included in A/I publications. Discipline-oriented A/I services could also collect difficult-to-obtain documents and, for those not limited by copyright, give them to a repository, noting the repository accession number in the printed A/I journal.

Personnel directories, subscription and membership files, also appear fruitful sources of information. It appears that they can serve three main purposes: (a) They hopefully provide accurate information on addresses, of immediate concern in publishing local directories, etc. (b) They provide a basis for improving current author indexes. Most author indexes are *name* rather than *author* indexes, since publications only list authors' initials. From affiliation and other supplementary data, it would be possible to more accurately identify an author. (c) They often contain information on individuals' subject matter interests. The potential utility of such information for SDI purposes is obvious.

Finally, I would like to suggest a procedure for adding a major information item to A/I files, namely, citation information. At present, the addition of such information is costly due to (a) lack of standardization in bibliographic format with the result that a variety of formats must be normalized, and (b) the character space required for references. The suggestion is that A/I publications offer a service to editors of selected primary journals to check references given in manuscripts against their cumulative files. For references for which abstracts have been published, the A/I service would furnish the abstract accession number that would then be printed as part of the citation in the primary publication. The editor of the primary journal has thereby been assisted through a check of the cited references. (This is not unimportant, since a substantial number of references appear to be inaccurate.) The cost to the primary journal is the inclusion of abstract accession numbers in the printed references. The reader of the primary journal will know immediately how to locate an abstract of the reference. One clerk is likely to be able to provide reference checks for 10 to 15 journals. An estimate for items in *Psychological Abstracts* indicates that this procedure would permit coverage of 80 to 85 percent of the journal articles cited, 50 percent of the books, and 25 percent of the technical reports. Gaps in prior coverage could also be identified through this procedure.

It is somewhat difficult to anticipate all of the potential uses of citation information. It is evident, however, that citation patterns provide an ability to discriminate the active from the inactive literature.

CHARLES H. STEVENS

Massachusetts Institute of Technology
Project Intrex
Cambridge, Mass.

IMPROVING USER ACCESS TO INFORMATION THROUGH APPLICATIONS OF TECHNOLOGY[1]

MIT's Project Intrex is an experimental, pilot-model, machine-oriented library system. The system includes a computer-stored catalog of 10,000 journal articles in selected fields of materials science and engineering, and the full text of the 10,000 articles on microfiche. The catalog is contained in a general-purpose, time-shared computer and is accessed through specially designed alphanumeric consoles, one of which has been implemented to date. The consoles are connected to the central computer through a buffer/controller. The full-text microfiche collection is accessed through the alphanumeric console and the access facility is designed to provide guaranteed, rapid access to any document in the collection at locations that are remote to the store. Full text may be viewed, page by page, at the user's station by means of a separate storage-tube display, or a permanent copy may be obtained either on 35-mm. film or as an 8 1/2 x 11-in. print. The elapsed time from the ordering of a 35-mm. film copy to the availability of it at the film station is approximately 90 sec. The first page of an article appears on the storage-tube display within 7 sec. after an order is placed, and each succeeding page can be obtained within 3 sec.

Intrex will use this experimental library as a means of gaining insights into the design characteristics of large-scale operational systems of a similar kind. We are in the process of evaluating the merits of the system by making it available to a selected community of users who have a bona fide need for the information contained in the system, and to librarians who may wish to use it for reference purposes. It is our intention to alter the characteristics of the system as we learn about its strengths and weaknesses from our user community.

It should be noted that the system brings the library to the user; it circumvents his need to go to the library for the information he is seeking. To be acceptable as a working system, however, it must engender satisfaction from the viewpoints of completeness and relevance of the information retrieved; it must be easy to engage; and it must provide quick, reliable service at costs that are realistic.

Intrex is examining several issues with respect to in-depth cataloging and the extent to which in-depth cataloging is needed when

---

1/ This summary is a condensation of a paper by Professor J. F. Reintjes, "System Characteristics of Intrex," 1969.

guaranteed, rapid access to full text is, and is not, readily available at the user's station. As many as 52 different items of information are being entered for each journal article cataloged. Through monitoring of the frequency of requests for each item we shall be able to draw conclusions on the relative value of each item.

The possibility exists that easy access to full text may alter the behavior patterns of those using the catalog. Since some of our catalog information is frequently contained on the first page of the document itself (author, author's affiliation, journal name, volume number, page, abstract, and so forth), users have the option of obtaining this information either by retrieving it from the catalog or from the document itself. Our purpose is to investigate the factors which govern user behavior when seeking information of this kind.

For storage of full text we have chosen image storage on microfiche. Computer storage of full text was discarded because of the huge amount of storage required. The use of film ensures preservation of pictorial information, and microfiche is well-suited for journal-article-type literature. We are using COSATI standard microfiche with a reduction ratio of approximately 18 to 1 and sixty frames per microfiche. Since our documents are derived from the published literature, we store text in image form rather than digitally because of the ease with which page text can be converted to a microfilm image.

The full-text storage and scanning unit, called the central station, is time-shared by several receiving stations. In order to minimize the time each microfiche is out-of-storage, each frame on a microfiche is scanned only once by means of a flying-spot scanner. Video signals are transmitted as analog information and to each frame of transmission is added a receiving station address in digitally encoded form. Single-frame transmission requires storage at the receiving stations; this storage is photographic in the case of the film station and electronic in the case of the storage-tube receiving station.

Our investigations show that at least 2,000 scan lines are required to reproduce the stored images with acceptable legibility. Variations in type sizes and quality of printing among documents dictate this scan-line requirement; in fact, as many as 3,000 lines may be needed where the quality of print is marginal and small type size is present, as in the superscripts and subscripts of mathematical equations.

The alphanumeric-console system through which the catalog is accessed has been developed as an experimental vehicle to determine special attributes which should be included for user convenience in library applications and to investigate techniques which might be employed to minimize cost of such consoles. The salient features of the catalog console system are these: typewriter keyboard input; refreshed CRT-display output; and a buffer/controller, consisting of a 128-track drum storage device and a Varian Data Machines 6201 computer, interposed between the central computer and the display consoles. The buffer/controller is capable of serving a maximum of ten display terminals. In addition to the keyboard, a set of programmed and programmable buttons is being provided in an effort to determine the usefulness of this approach as a user aid.

In the original system design, catalog information and full text are displayed on separate CRT's. This configuration is obviously inconvenient and costly; a single display is a goal of our project.

A critical item with respect to utilization of a machine-stored library is the amount of preparation needed to engage the system and to make full use of its power. The bona fide user is interested only in satisfying his need for information as completely and quickly as possible, and with a minimum amount of preliminary effort. Nevertheless certain procedural matters must be mastered even before intelligible responses can be derived from the machine. These include an understanding of the options the user has for making searches and typing procedures for executing these searches, as well as an ability to type and to invoke proper procedures for correcting typing errors. Since many library users are occasional or intermittent users, they may always be in the category of *new* users in the sense that their retentivity of basic operational rules from system engagement to system engagement may be minimal. User aids thus become a crucial item.

Thus far, Intrex has experimented with several types of user aids. A User Guide which describes in detail the various features of the system and how to use them is available both *off-line* in hard-copy form, and *on-line* as a computer printout. In addition, the Guide is available in summary form in a separate booklet entitled *How to Get Started*. Wall charts describing system operation are also posted directly before the user above his console. Still to be prepared and tested are simple instructions on cards.

The results of our experiments to date indicate that the off-line Guide is the version referred to most frequently; apparently, the time required to print out the various sections of the on-line version is considered to be wasteful. Nevertheless dissatisfaction with the off-line Guide has been expressed. Dissatisfaction seems to result from the large quantity of material it contains, the time required to assimilate the material, and the fact that the phraseology includes a certain amount of technical jargon which is not understood. A further observation is that a summary version of the Guide is helpful, but it must be supplemented by the full-scale version. Finally, response to wall charts as user aids has been disappointing thus far. Further experimentation with their content is needed in order to determine their value as a user aid, if indeed they have value.

To date experiments have been conducted only with a typewriter console as an input-output device. A new dimension to user-aids is added when a graphical terminal becomes available. Since its writing rate is an order of magnitude faster than that of the typewriter, current disinterest in the on-line version of the instructional guide may disappear. Furthermore, since the Intrex graphical console is being designed with a set of programmable switches, these may offer a wholly new approach to on-line instruction when these switches are programmed as user aids.

ROWENA W. SWANSON

Air Force Office of Scientific Research
Arlington, Va.

PROPOSALS FOR ENHANCING THE HANDLING OF INFORMATION

Though the "publish or perish" credo may not have been an overt driving force three and more decades ago, publication has always been a modus operandi for the person who discovers, or thinks he has discovered, new knowledge and wants to share it with his fellow man, be this drive for monetary gain, self gratification, or more selfless purposes.

Reasonably generous support of the scientific and academic communities over the past two decades, together with increasing applicability of a new tool, the computer, to information processing have perhaps more than additively combined to create an impressively large volume of current publication and exponentially increasing growth rates. The volume itself is having somewhat of a sobering effect and the computer gives evidence of some labor-saving assistance, but neither offers the promise of stemming the rise in the growth of quantities of information that are being generated today, and will continue to be generated.

It should be noted that the generation of information, per se, is scientifically desirable. It is through new knowledge--discovered, pondered, digested, interrelated with other knowledge, some new and some old--that science has contributed so remarkably to the medical and technological states of well being our society enjoys today.

It is, in fact, because we consider information so valuable that we don't want to lose it that we--both the producers and users of information--have been trying to develop innovative ways of *packaging* or containing and presenting information, ways of compressing and digesting it to keep it within the human overload threshold, ways of selectively disseminating and communicating it, etc.

One alternative to the information problem is, of course, to stop it at its source, that is, to halt or slow down its production. This obviously would impair goals for our society that anticipate ever-greater incorporation of science-based advances into the working and personal lives of our citizens. It must be realized, however, that loss of information once it has been produced is at least as wasteful, if not more so, than not having produced it at all.

Imagination has been at work over the past twenty years in the handling of information. We have seen the rise of secondary publications, principally the useful abstract periodical with its associated indexes, and the mushrooming of information centers that have been functioning as libraries for scientific materials and as services offering bibliographic, reference, and sometimes more extensive aids to the inquirer. Publishers have stimulated the preparation of many monographs that have tended to keep reasonably current with new information, and the critical review and the annual review have grown in popularity and recognized importance for their tutorial and summarizing values. Less formal communication has also been on the upsurge. The shortcuts to information transfer attendant in corridor and off-the-cuff discussion have heightened the significance of the meetings of many professional societies. Special-interest groups, some formally structured and others more or less ad hoc, have also exploited the conference device (usually restricting the attendance) as well as exchanges for informal publications (newsletters, correspondence, pre-publication papers, etc.).

It would be unjust to ignore the contributions that all of the primary and secondary publishers and producer and user groups have made. These contributions make the present scene crowded but not chaotic. One can, albeit laboriously, locate journals, papers, data, persons responsible for particular ideas. One can even not know who or what one is looking for and find some measure of satisfaction in a response if one is willing to wander through the existing people/publications complex long enough. This, it is contended, is a better-than-nothing situation, and it often works reasonably efficiently. It does, however, rely too much on the inquirer's patience, his knowledge of the people/publications complex, and the patience and good will of those whom he interrogates as well as their own knowledge of the complex.

Is the imposition of a high degree of order within information transmission channels and on their interrelationships the basis for a solution to all information handling problems? It is believed that better order would, indeed, at least partially alleviate some difficulties that presently exist in locating information sources, understanding scopes and overlaps and areas of responsibility, etc. It is also believed that better order might tend to identify overlaps and gaps in coverages and services and might lead to mutually reinforcing cooperative arrangements that do not now exist. The source of the order could be a seat of guidance that has both an overview and an objectivity and that could also perhaps function as an arbiter.

What, then, must accompany better order? It is suggested that a more realistic appraisal is needed of not only the complex and the things (computers, journals, libraries, services) in the complex, but also of the people in the complex, all of the people. It is contended, for example, that people who want information are not all the same, that is, they do not all behave in the same way. It is obvious, of course, that the financial manager and the technical director and the bench scientist are likely to want different items of information that could all pertain to the same entity in a more generic sense. It is less obvious but can be demonstrated that two bench scientists or two financial managers might also want items of information packaged in different ways in accordance with their own idiosyncratic preferences for processing information. Can one say, idiosyncratic preferences, humbug? Perhaps. But which are humbug, and which might be importantly significant to the scope of the processor's enterprise, that is, which had better be paid attention to because they are important to the man's productivity? Different behaviors, different purposes, different information needs, different time priorities, different levels of experience, different geographic locations--all these and more bear on what information may be appropriate for a particular person at a particular time in a particular place. Automation may be able to accommodate some of these characteristics, people and machines may be needed for others. The mixes and matches have to be determined, and their determination is part of today's exciting frontier toward more effective use of information tomorrow.

CHARLES W. N. THOMPSON

Northwestern University
Department of Industrial Engineering & Management Sciences
Evanston, Ill.

EXPERIMENTS WITH RESPONSES OF USERS: PRELIMINARY FINDINGS

A few background comments may serve to introduce the remarks which will be made later in this paper. The Program of Research on the Management of Research and Development at Northwestern is in its tenth year; it existed for four years prior to this at MIT, and had its inception at Columbia U. The program is an integral part of the organization theory area of the Department of Industrial Engineering and Management Sciences, and the training and experience of our faculty and advanced graduate students include a wide variety of disciplines within and outside of engineering. Our major approach is one of the development of fairly formal propositions followed by testing under field conditions in operating organizations. In the past five or six years, we have increasingly used field experiments in addition to study, and our long interest in the flow of ideas and the communication process has provided a basis for several detailed studies of the information-related behavior of scientists and engineers.

One of our earliest attempts to experiment with the information environment in an operating laboratory involved 18 X-ray crystallographers in nine industrial and nonindustrial laboratories in the Chicago area. After a short period of observation, we provided them with seven services for another short period, then extended the services for a period of six months. Without going into the experiment in detail, and noting that the subjects were otherwise most cooperative, the point is that they did not respond to their new services by making permanent, observable changes in their searching behavior.

In two related experiments, this time with medical researchers, we provided the services of an experienced medical librarian. In the first, and smaller, experiment, we found some (at least temporary) changes in behavior. In the larger experiment, involving 12 groups of medical researchers, and starting out with 110 participants, we provided both the librarian and facsimile equipment to six of the groups for 3 months, and then switched the service to the other six groups, observing both groups during both periods. There were, as might be expected, considerable variations in the frequency and characteristics of the usage, and we attempted to relate various personal, professional, job, and organizational characteristics to each individual's use or nonuse of the system.

Finally, two studies involved the actual use of documents. The first studied the *usefulness* of technical reports and literature arriving on the desk of a group of researchers; one of the things observed was that a researcher will call an item *useful* even if it does not cause him to take some action. The second, and larger, study was an *administrative experiment* involving 85 scientists and engineers at three government laboratories. Here, the purpose was to discover whether the presence of an abstract accompanying a document was related to shorter times and *better* decisions with respect to the initial disposition of the document, and the result was that there appeared to be no significant difference.

We are involved in other investigations, one a longitudinal study to trace changes in *style* over a period of five years, but one of the points which our work suggests, and which we find in studying the research of other investigators, is that scientists and engineers do not respond to the mechanism--the systems and services that are provided--in neat, stable, easily categorized ways.

Commentary

In the face of the voluminous research literature on the information process and the many programs to develop and test new information systems and services, it may seem difficult to suggest that more needs to be done. One of the things that we can do is to improve our own information process by some of the many devices we already have available, including more comprehensive reviews and citations and previous work in our own reports, assuring that reference is made to where the complete report of our research and supporting data may be found, and cooperating with those who are attempting to *organize* and summarize the state of the art.

That which seems self-evident is that the information process --sources, transfer mechanism, users--is pigheadedly pluralistic. If we are to make progress, we must recognize that the solutions we propose are, and should be, for the most part modest modifications of processes that are already in use in some form somewhere. One of these areas is personal files. User studies point out how ubiquitous they are. These vary from a modest stack in a corner to elaborate, machine-indexed systems. There are, for example, a number of systems among those working on our Northwestern program--alphabetic files indexed with 3 x 5 cards, serial files with computer-printed author-title indexes, title and abstract computerized files, etc. Further afield, individual and small group filing systems are endless, some competing with or even drawing on library systems. In all of these, we can find the expression of a real need for help. One obvious possibility is to find that minimum common denominator that would provide the basis for a degree of what in other areas is called standardization. Some minimum compatibility, some basis for a transform that might save those who are still trying to develop personal filing systems the effort, and allow those who have filing systems to draw on one another, to communicate, to exchange, to build and, perhaps, to take advantage of the economies of common machine programs and information stores. In our program, we are working with a *minimum identification* system along these lines for our field data.

This argument for a minimum compatibility stems from a recognition that even our largest and most puissant systems--DDC, Library of Congress, or the major journals--have not only not brought conformity but have difficulty achieving recognition and acceptance. Perhaps it is in order to spend more time exploring systems that do a minimum for the individual, that provide him with an opportunity to react to the system over the range from nominal awareness and the minimum identification of the document all the way to the richest and most variegated interaction with all the various forms, including the whole document (and perhaps its gestalt--the author's vitae, the refereeing process of the publisher, etc.).

It would seem that there are a number of relatively simple and mundane things which could be done to improve the process. Most of these are already being done, but unevenly. For example, in the era of the photocopier, a full citation printed on the first page of each journal not only saves time, but it also avoids the later *look up* and the too-frequent erroneous citation. Similarly, the title page or the overleaf on reports could contain a standard citation, especially to provide a clue to handling indicia like "Final Report ..." and the multi-line designation of the publishing organization. None of these present problems to the professional librarians, but they do inhibit the individual. Similar comments apply to the title pages of books.

From our study of the use of abstracts, mentioned above, we learned that the scientists and engineers did not find themselves overburdened by having to handle documents that arrived at their desks. We also found that they drew on many sorts of cues, both within and outside the immediate document, to determine whether the document was useful or not. They tended to read, or retain for reading, quite a few of the documents. It seems probable that their behavior might well result in physical storage problems, and we did not attempt to determine whether they might have a retrieval problem. If, however, the *presence* of a document does not present an information overload, if the individual can deal with the sorting problem on terms at least acceptable to him, it may suggest that mechanisms, such as the interactive display, or the microfiche, that would bring many documents quickly and easily to him, will be of central importance. And if we make a record of his interactions, and especially the patterns of his requests, we may be able to establish priorities in the development and implementation of the facilities, such as buffer libraries, various surrogates, and decision trees.

In summary, these comments suggest that broad palliatives, or detailed formal systems may be inherently unproductive, at least until we learn far more about the process by which the individual determines what he wants and how he wants it. It is suggested that the variety of *systems* that people use may provide clues to those minimum systems that will be most adaptive, and most used.

DAVID WEEKS

George Washington University
Biological Sciences Communication Project
Washington, D. C.

THE ROLE OF INFORMAL COMMUNICATION

A study in progress at BSCP deals with patterns of exchange of information through means other than journals. This area appears less readily amenable to technical solutions. Here we are concerned with direct and indirect interchange between individuals or a speaker and his hearers. This form of communication is probably the most frequently used; it deals mainly with smaller units of information (conversations and correspondence) and it is much more difficult to measure in quantitative terms. It is also a convenient refuge for respondents to surveys who prefer not to record a reliance on specific sources of information.

Research on informal communication has been primarily descriptive. Some of the data available come from user studies made for other purposes. Little is known about the amount of information and how much of that amount is new, which actually passes from one scientist to another through informal means. Perhaps because of its amorphous nature and the facile expression *invisible college* the whole subject has tended to assume dimensions which may be misleading and unlikely to stand very close scrutiny.

One illustration of an attempt to set in motion an informal exchange and to observe its activity was the experiment sponsored by the National Institutes of Health. The experiment of an Information Exchange Group, which lasted from 1962 to 1966, was extended to cover seven select groups of scientists. The IEG was indeed a novel concept, sufficiently costly to require a government sponsor with facilities and resources to support a substantial effort. Unlike studies which have sought to find generalizations from observing the real world, the NIH program was a true experiment in which the individuals who participated were formally engaged in the effort and the information exchanged was itself a part of the experiment.

Briefly, the concept was that the mechanism of information exchange would focus on a specific research topic and gather the relevant *invisible college* into group membership. In all, seven such groups were formed ranging in size from 250 to 1500 members. Group number one came closest to representing the original idea--its topic was Oxidative Phosphorylation and Terminal Electron Transport. Even this topic brought 750 members, either nominated by a member or admitted at their own request. But by the time a seventh group was formed, the subject scope had broadened almost to a generality--Nucleic Acids and the Genetic Code. Altogether, 3600 scientists participated and of the original seven groups, four are still in existence but operate outside the framework of NIH sponsorship.

Clearly, there are some conclusions that can be derived from this experience, yet even these are exceedingly weak on questions relating to why scientists' information gathering habits follow any particular pattern. The experiment, however, was not conducted to yield answers to this question. Instead, it was aimed at trying out a vehicle for conveying substantive information and a mechanism for handling the information. In both respects it accomplished its purpose, but our interest now is in using the record of what took place as a means of getting insights into other aspects of information exchange.

One obstacle to the collection of definitive data on the whole subject of communication is the unreliability of any form of questionnaire or interview. Even the diary technique used by the American Psychological Association is not free from distortion. The preparatory studies for the formation of the Lister Hill Biomedical Community Network point out the dangers of poll-taking in a community characterized by a diversity of interest and sophistication. The System Development Corporation in approaching the task of evaluating information systems started with recognition of the fact that judgments of relevance were valid only when correlated with the users' purpose in seeking information.

This same problem, which assumes a somewhat different form in studies of information media (journals, preprints, papers, correspondence), shows up clearly in responses to different inquiries. It would be necessary to assemble the responses to a considerable series of studies to demonstrate the inconsistencies which seem to characterize their results. While such a hypothesis may prove false, it is possible that the apparent purpose of a survey has affected the responses to a degree that makes much of the data questionable. The kind of anomalies that may appear suggest that respondents may tend to record their preferences with two objectives: 1) to place themselves in the best possible light with regard to keeping current to their special field, 2) to avoid a firm commitment to reliance on the information medium under study in the questionnaire or interview.

These problems are perhaps more properly dealt with in studies of the sociological or psychological aspects of scientific communication. But it is reasonable to assume that inconsistent findings are in need of explanation before poll-taking data can be considered adequate for decisions about improved mechanisms for information exchange. These inconsistencies show up in different kinds of studies. When a community of document system users (or potential users) is questioned, the results show that journals are a major source of information; when journal utilization is studied, a strong reliance on advance information from preprints and manuscripts is reported. When the Information Exchange Groups (NIH) membership was questioned about the preprint exchange, many indicated a preference for receiving only titles or abstracts on the single topic which was the precise focus of group membership.

When the results of surveys produce so little confirming evidence, it can scarcely be said that we know very much about the uses of literature and information in the scientific communities. One of the difficulties clearly lies in the tendency of aggregate numbers to merge different segments of a community into a whole and to conceal their differing interests. Studies addressed to a discipline are especially vulnerable. Even mission-oriented groups are composed of smaller entities distinguished by their functions. But an awareness of the many different needs and uses cannot alter the fact that both publication and information handling systems require large user communities in order to maintain their activities and are limited in their ability to address different services to segments of those communities.

Since multiple services addressing the preferences of smaller and smaller interest groups is clearly too costly to become the likely direction of future effort, it is important to be certain that general purpose services are indeed based on a clear perception of needs and patterns of recourse to information. The same is true for the scientific communities whose preferences and habits are being studied. Their own perceptions may need revision if any innovations in the communication process are to be developed. Consider briefly some of the findings from two studies. The first was made in 1969 when the Entomological Society of America was canvassed to collect data about the present use patterns and the media employed maintaining current awareness in the field.

In response to a question about utilization of sources, 743 respondents out of a total of 2334 reported extensive use of abstracting journals. Yet 1414 reported that they regularly utilized *Biological Abstracts*. The most often cited source extensively used to obtain information was *professional journals* (1740 respondents) representing 74.8 percent of the total. This suggests a significant reliance on journals, particularly when compared with the 31 percent who reported extensive use of abstracting journals. More than three fourths of all respondents expressed satisfaction with the sources available to them--both primary and secondary--and did not feel a lack of access to be a problem. This would be a gratifying response if it did not contrast with the scope and coverage of primary literature reported under the heading of core material. Of the more than 850 journals devoted to Entomology in the world's current literature, only 147 were reported as sources by the respondents. Either the community are unaware of the major portion of the journals devoted to this subject or that same portion of the literature (75 percent) is so lacking in new or significant research material as to be unnecessary for publication.

When we turn to the *invisible college* as the medium for exchange, bypassing the journal and its time-lag problem, another study offers some curious findings. Admittedly an informal and unpublished study, still the outcome casts reasonable doubt on the role of this mechanism. Here, each scientist was asked to name the authorities in the field; the individuals names were then asked to make a similar list. It was hoped by this means to discover enough commonality to gain some insight into an *invisible college*. Instead, the results showed no overlap at the first level and almost none at the second.

Overall, these illustrations have been drawn to point to the anomalies both within and among efforts to study the exchange and use of information. A definitive comparison of many pieces of study data, like the user studies themselves, would probably lead only to predictable results. We can also venture some conclusions which can have an impact on the direction of efforts aimed at special solutions to special information problems. 1) The results of extensive research into scientists' literature and information utilization practices are too imbalanced to serve as a basis for generalization useful in systems development. 2) If there is any real benefit in seeking more definitive information, then new methods and techniques are necessary. 3) The chief problem may lie in the momentum of publication, and if so, qualitative studies and some self restraint may achieve more than searching for technical solutions in the production of literature.